10:23:44 OCA PAD INITIATION - PROJECT HEADER INFORMATION 09/27/91

Active

Project #: E-21-655 Cost share #: Rev #: 0
Center # : 10/24-6-R7325-0A0 Center shr #: OCA file #:
Work type : RES
Contract#: N00014-91-J-4129 Mod #: Document : GRANT
Prime #: Contract entity: GTRC

Subprojects ? : N CFDA: 12.AAA
Main project #: PE #: N/A

Project unit: ELEC ENGR Unit code: 02.010.118
Project director(s):
WILLIAMS D B ELEC ENGR (404)853-9832

Sponsor/division names: NAVY / OFC OF NAVAL RESEARCH
Sponsor/division codes: 103 / 025

Award period: 910801 to 920731 (performance) 920731 (reports)

Sponsor amount	New this change	Total to date
Contract value	82,500.00	82,500.00
Funded	82,500.00	82,500.00
Cost sharing amount		0.00

Does subcontracting plan apply ?: N

Title: TRANSFORM-BASED WIDEBAND ARRAY PROCESSING

PROJECT ADMINISTRATION DATA

OCA contact: E. Faith Gleason 894-4820

Sponsor technical contact

RABINDER N. MADIN
(703)696-4217

OFFICE OF NAVAL RESEARCH
CODE 1114SE
800 N. QUINCY STREET
ARLINGTON, VA 22217-5000

Sponsor issuing office

EDNA MCHALE
(703)696-2605

ONR
CODE 1512B:EAM
800 N. QUINCY STREET
ARLINGTON, VA 22217-5000

Security class (U,C,S,TS) : U ONR resident rep. is ACO (Y/N): Y
Defense priority rating : N/A ONR supplemental sheet
Equipment title vests with: Sponsor GIT X

Administrative comments -
INITIATION.

GEORGIA INSTITUTE OF TECHNOLOGY
OFFICE OF CONTRACT ADMINISTRATION

NOTICE OF PROJECT CLOSEOUT

Closeout Notice Date 08/24/92

Project No. E-21-655 Center No. 10/24-6-R7325-0A0

Project Director WILLIAMS D B School/Lab ELEC ENGR

Sponsor NAVY/OFC OF NAVAL RESEARCH

Contract/Grant No. N00014-91-J-4129 Contract Entity GTRC

Prime Contract No. _____

Title TRANSFORM-BASED WIDEBAND ARRAY PROCESSING

Effective Completion Date 920731 (Performance) 920731 (Reports)

Closeout Actions Required:	Y/N	Date Submitted
Final Invoice or Copy of Final Invoice	Y	
Final Report of Inventions and/or Subcontracts	Y	
Government Property Inventory & Related Certificate	Y	
Classified Material Certificate	N	
Release and Assignment	N	
Other _____	N	

Comments _____

Subproject Under Main Project No. _____

Continues Project No. _____

Distribution Required:

Project Director	Y
Administrative Network Representative	Y
GTRI Accounting/Grants and Contracts	Y
Procurement/Supply Services	Y
Research Property Managment	Y
Research Security Services	N
Reports Coordinator (OCA)	Y
GTRC	Y
Project File	Y
Other _____	N
_____	N

NOTE: Final Patent Questionnaire sent to PDPI.

Semi-Annual Performance Report
August 1, 1991 — January 31, 1992

Transform-Based Wideband Array Processing

Douglas B. Williams, Principal Investigator
Rabinder N. Madan, Scientific Officer

Contract No.: N00014-91-J-4129
R&T No.: 4148136- - -01

Over the past six months, this contract has funded two projects in full and one project in part. The two fully funded projects focus on the application of random coefficient models to wideband high-resolution direction finding and transient signal detection and estimation. The partially funded project involves the analysis of nonlinear, possibly chaotic, dynamical systems. Progress in each of these areas is described below.

Random Coefficient Models: The application of random coefficient models to narrowband high resolution direction finding has been very successful and has already resulted in a conference paper to be presented at ICASSP92 (see the attached paper by Jost & Williams). It has been shown that the random coefficient model is much better suited to modeling sensor array data than the autoregressive model is. A simple method for estimating the parameters of the random coefficient model has been developed and applied to simulated data. Finally, a beamformer for the random coefficient model has been developed which has significantly better performance then earlier linear predictive beamformers. A journal paper describing these results is currently being written and should be submitted to the *IEEE Transactions on Signal Processing* by this Spring. There is also progress being made toward the final goal of applying these techniques to wideband direction finding.

Transient Signal Detection: The application of the wavelet transform to the detection of transient signals with an array of sensors is being examined. This approach has led to a directional multirate filter bank structure that decomposes the incoming signal into decaying exponentials. This filter bank is also capable of adapting towards an improved estimate of the structure of the transient signal which, consequently, also improves the detection performance. Publication of these results will proceed after simulations and comparisons to other transient signal detectors are complete.

Nonlinear System Identification: System identification algorithms that depend on gradient descent methods have been found to degrade significantly if the time-series or, equivalently, the system that produced the time-series is chaotic (see the attached paper by Drake & Williams). A careful analysis of these degradations has led to algorithms which are much less sensitive to the potentially chaotic nature of these nonlinear systems. Analysis of these systems and their time-series continues with the eventual goal being a very general system identification algorithm to be applied to the time-series produced by nonlinear discrete-time systems.

To be presented at ICASSP, San Francisco, California, March 1992.

A BEAMFORMER BASED UPON THE RANDOM COEFFICIENT MODEL *

Bruce Jost *Douglas B. Williams*
Georgia Institute of Technology, Atlanta, GA, 30332-0250 USA

ABSTRACT

This paper applies the random coefficient model to array processing, specifically in the design of a beamformer for direction finding. This model is similar to the autoregressive (AR) model, except the coefficients are allowed to change with time instead of remaining constant; thus allowing the beamformer to better model any additive noise or signal correlations in the observations. Through the use of a binary hypothesis test, it is shown that random coefficient models better fit typical array data than do AR models. A Kalman filter is presented that has the array observations as inputs and the parameters of the random coefficient model as outputs. A new beamformer based on the random coefficient model is derived that is similar to the constant coefficient linear predictive (CCLP) beamformer. The two beamformers are compared and it is shown that the random coefficient beamformer outperforms the CCLP beamformer.

1. INTRODUCTION

The linear model is probably the most popular model in use in engineering and science because it is simple and yet powerful. Many times, computational advantages due to the simplicity of the linear model far outweigh any performance gain achieved by more complicated models. The commonly used autoregressive (AR) model is one such linear model. In array processing the AR model leads to linear predictive (LP) beamforming [1] where the output of a selected sensor, say the m_0th, is estimated as a weighted linear combination of the other sensor outputs. Assuming narrowband signals and letting $X_m(k), k = 0, \ldots, N-1$, be snapshots of the Fourier transform of the mth sensor's output at the signals' common center frequency, the LP model for $X_{m_0}(k)$ is

$$X_{m_0}(k) = -\sum_{m \neq m_0} a_m X_m(k), \qquad (1)$$

where $\{a_m\}$ is a set of complex-valued weights to be found.

When compared to other beamformers with roughly equivalent computational complexity, such as the minimum variance distortionless receiver, the LP method possesses superior resolution properties [1]. However, the performance of the LP beamformer is highly dependent on both the signals' directions of propagation and the selection of m_0. Also, in many cases the direction-of-arrival estimation bias is quite high. The primary causes of this estimation bias are additive noise in the observations and correlations between propagating signals. Additive noise and signal correlations do not obey an AR relationship across the array; so the model in (1) is no longer accurate.

In order to keep the simplicity of the linear model of (1) and yet more accurately match the array data, the random coefficient model is proposed. This model is identical to the AR model except the weights a_m are random instead of constant and (1) becomes

$$X_{m_0}(k) = -\sum_{m \neq m_0} a_m(k) X_m(k). \qquad (2)$$

The coefficients can be expressed as

$$a_m(k) = \beta_m + \nu_m(k), \qquad (3)$$

where β_m is the mean of the coefficients for the mth sensor and the $\nu_m(k)$ are zero-mean independent identically distributed random variables. The random coefficient model of (2) keeps the simple linear form of the AR model and, hence, its computational advantages. Since the constant coefficients of an AR model can be expressed in terms of (3) with the $\nu_m(k)$ equal to zero, the random coefficient model can actually be thought of as a generalization of the AR model. Therefore, the random coefficient model will fit the array data at least as well as the AR model and will simplify to the AR model when it is the better choice.

2. TESTING OF THE MODELS

Through the use of a binary hypothesis test developed by Breusch and Pagan [2], it is possible to test which model, AR or random coefficient, will better fit typical array data. The test indicates that if the residuals from an AR model have a constant variance, an AR model is better suited; otherwise, a random coefficient model should be considered. The test is performed by generating a Lagrange multiplier (LM) statistic from generated array data. If the null hypothesis of an AR model is true, then the LM statistics will be asymptotically distributed as chi-squared with M degrees of freedom, where M is the number of sensors in the array. The method used for this paper was to generate

*THIS WORK WAS SUPPORTED BY THE OFFICE OF NAVAL RESEARCH UNDER CONTRACT NUMBER N00014-91-J-4129

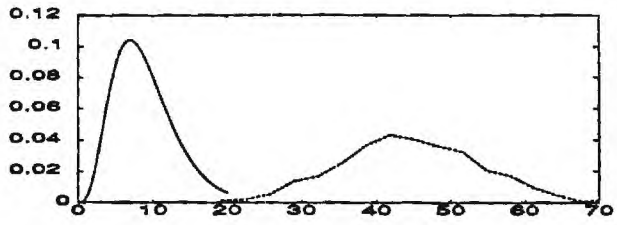

Figure 1: Results of the hypothesis test. The ten degree of freedom chi–squared distribution is the solid line while the normalized histogram of the LM statistics is the broken line. The array data was from a uniform linear array of 10 sensors with spacing $\frac{\lambda}{2}$. There are two signals present at $-10°$ and $30°$, the noise is additive white Gaussian and the SNR is 6 dB. There are 100 independent samples being used ($N = 100$).

array data for a given number of signals at their respective angles of arrival, and a given number of sensors, time samples, and signal-to-noise ratio. The LM statistic given in [2] was calculated for a large number of sets of data and the normalized histogram of these LM statistics was calculated and plotted against a chi–squared density function with M degrees of freedom to see how closely the two matched. Figure 1 shows an example of this test in which the normalized histogram of the LM statistics clearly does not match the chi–squared density plot. In fact, it was found that the AR hypothesis is rejected for even the simplest cases if any additive noise or signal correlation is present. Therefore, it would seem that the random coefficient model is better suited for array processing applications.

3. RANDOM COEFFICIENT BEAMFORMER

As the random coefficient model has been shown to fit the data more accurately than the AR model, a beamforming algorithm based on this model has been developed. The remainder of the paper will present the algorithm for estimating the random coefficients and then the derivation of the random coefficient beamformer.

3.1. Estimating the Coefficients

The first step is to estimate the random coefficients. This is performed with the Kalman filter outlined in [3] and [4]. The following set of vectors is introduced to ease the computations:

$$\begin{aligned}
\mathbf{a}'(k) &= [a_0(k)\cdots a_{m_0-1}(k)\, a_{m_0+1}(k)\cdots a_{M-1}(k)] \\
\mathbf{b}' &= [\beta_0\cdots\beta_{m_0-1}\,\beta_{m_0+1}\cdots\beta_{M-1}] \\
\mathbf{v}'(k) &= [\nu_0(k)\cdots\nu_{m_0-1}(k)\,\nu_{m_0+1}(k)\cdots\nu_{M-1}(k)] \\
\mathbf{z}'(k) &= [X_0(k)\cdots X_{m_0-1}(k)\,X_{m_0+1}(k)\cdots X_{M-1}(k)]
\end{aligned}$$

where M is the number of sensors and ($'$) denotes Hermitian transpose. An error covariance matrix is defined as

$$\Sigma = [\sigma_{ij}^2],\ i,j = 0,\ldots,m_0-1,m_0+1,\ldots,M-1$$

where $E\{\nu_i(k)\nu_j^*(l)\} = \delta(k-l)\sigma_{ij}^2$. From [3] and [4] the estimates of $\mathbf{a}(k)$ are calculated from the Kalman filter equations

$$\begin{aligned}
\hat{\mathbf{a}}(k/k-1) &= \mathbf{b} \\
\hat{\Sigma}(k/k-1) &= \Sigma \\
h(k) &= \mathbf{z}'(k)\hat{\Sigma}(k/k-1)\mathbf{z}(k) + \sigma_n^2 \\
\hat{\mathbf{a}}(k/k) &= \mathbf{b} + \hat{\Sigma}(k/k-1)\mathbf{z}(k)h^{-1}(k) \\
&\quad \cdot [X_{m_0} + \mathbf{z}'(k)\mathbf{a}(k/k-1)] \\
\hat{\Sigma}(k/k) &= \hat{\Sigma}(k/k-1) - \\
&\quad \hat{\Sigma}(k/k-1)\mathbf{z}(k)h^{-1}(k)\mathbf{z}'(k)\hat{\Sigma}(k/k-1) \quad (4)
\end{aligned}$$

where $\hat{\mathbf{a}}(k/k)$ is the estimate of $\mathbf{a}(k)$ at time k given all the observations up to k, $\hat{\Sigma}(k/k)$ is the error covariance matrix estimate of $\mathbf{a}(k)$ at time k given all the observations up to k, and σ_n^2 is the noise variance estimate.

To implement the Kalman filter, initial estimates of $\mathbf{b}, \Sigma$, and σ_n^2 must be calculated. The authors will present a method that calculates the least–squares estimates of these terms following [4]. Introducing a noise term into (2) yields

$$X_{m_0}(k) = -\sum_{m\neq m_0} a_m(k)X_m(k) + n_{m_0}(k) \qquad (5)$$

where n_{m_0} is white Gaussian noise and is uncorrelated with the observations $X_m(k)$. The noise associated with an arbitrary sensor, say the mth, has the property $E\{n_m(j)n_m^*(k)\} = \delta(j-k)\sigma_n^2$. Using (3), equation (5) can be written as

$$\begin{aligned}
X_{m_0}(k) &= -\sum_{m\neq m_0}\beta_m X_m(k) - \sum_{m\neq m_0}\nu_m(k)X_m(k) \\
&\qquad + n_{m_0}(k) \\
&= -\sum_{m\neq m_0}\beta_m X_m(k) + u(k) \qquad (6)
\end{aligned}$$

where

$$u(k) = n_{m_0}(k) - \sum_{m\neq m_0}\nu_m(k)X_m(k). \qquad (7)$$

Defining

$$\begin{aligned}
\mathbf{u}' &= [u(0)\cdots u(N-1)] \\
\mathbf{y}' &= [X_{m_0}(0)\cdots X_{m_0}(N-1)] \\
\mathbf{Z}' &= [\mathbf{z}(0)\cdots\mathbf{z}(N-1)],
\end{aligned}$$

equation (6) can be written as

$$\mathbf{y} = -\mathbf{Z}\mathbf{b} + \mathbf{u}.$$

Using ordinary least squares, the estimate of $\mathbf{b}$ is

$$\hat{\mathbf{b}} = -(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}. \qquad (8)$$

Taking the expected value of the magnitude squared of (7) yields

$$E\{|u(k)|^2\} = E\{|\, n_{m_0}(k) - \sum_{m\neq m_0}\nu_m(k)X_m(k)\,|^2\}$$

which reduces to

$$E\{|u(k)|^2\} = \sigma_n^2 + \sum_{i\neq m_0}\sum_{j\neq m_0}E\{\nu_i(k)\nu_j^*(k)X_i(k)X_j^*(k)\} \qquad (9)$$

where the other terms are zero because the noise is uncorrelated with the signals and coefficients. Assuming the $X_i(k)$ and $a_i(k)$ are zero mean and Gaussian, (9) can be expressed as

$$E\{|u(k)|^2\} = \sigma_n^2 + \sum_{i\neq m_0}\sum_{j\neq m_0}[E\{\nu_i(k)\nu_j^*(k)\}E\{X_i(k)X_j^*(k)\} + E\{\nu_i(k)X_i(k)\}E\{\nu_j^*(k)X_j^*(k)\} + E\{\nu_i(k)X_j^*(k)\}E\{\nu_j^*(k)X_i^*(k)\}]$$

which reduces to

$$E\{|\,u(k)\,|^2\} = \sigma_n^2 + \sum_{i\neq m_0}\sum_{j\neq m_0} E\{\nu_i(k)\nu_j^*(k)\}E\{X_i(k)X_j^*(k)\}. \tag{10}$$

Since

$$|\,u(k+l)\,|^2 = E\{|\,u(k+l)\,|^2\} + \xi_l \tag{11}$$

where the ξ_l is a zero mean error with unknown variance, (10) can be used in (11) as

$$|u(k+l)|^2 = \sigma_n^2 + \sum_{i\neq m_0}\sum_{j\neq m_0} E\{\nu_i(k+l)\nu_j^*(k+l)\}E\{X_i(k+l)X_j^*(k+l)\} + \xi_l$$

or in vector notation as

$$\mathbf{e} = \mathbf{X}\mathbf{s} + \mathbf{r}$$

where $\mathbf{e}' = [|u(0)|^2 \cdots |u(N-1)|^2], \mathbf{r}' = [\xi_0 \cdots \xi_{N-1}], \mathbf{s}' = [\sigma_n^2\ \sigma_{1,1}^2\ \sigma_{1,2}^2 \cdots \sigma_{1,M-1}^2\ \sigma_{2,1}^2 \cdots \sigma_{M-1,M-1}^2]$, and

$$\mathbf{X} = \begin{bmatrix} 1 & X_1(0)X_1^*(0) & X_1(0)X_2^*(0) & \cdots & X_1(0)X_{M-1}^*(0) & X_2(0)X_1^*(0) & \cdots & X_{M-1}(0)X_{M-1}^*(0) \\ \vdots & \vdots & & & & & & \vdots \\ 1 & X_1(N-1)X_1^*(N-1) & & & & \cdots & & X_{M-1}(N-1)X_{M-1}^*(N-1) \end{bmatrix}$$

Using ordinary least squares, the estimate of s is

$$\hat{\mathbf{s}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}, \tag{12}$$

from which the estimates of σ_n^2 and Σ are extracted.

3.2. The Beamformer

Once the random coefficients have been estimated, they can be used in a beamformer to calculate the direction(s) of the measured signal(s). Following the derivation of the constant coefficient LP beamformer, (5) can be rewritten as

$$n_{m_0}(k) = X_{m_0}(k) + \sum_{m\neq m_0} a_m(k)X_m(k) = \sum_{m=0}^{M-1} a_m(k)X_m(k) \tag{13}$$

if $a_{m_0}(k) = 1$ for $k = 0, \ldots, N-1$. Multiplying (13) with a similar expression for $n_{m_0+l}^*(k)$ and taking the expected value gives (dropping the time index for brevity)

$$E\{n_{m_0}n_{m_0+l}^*\} = \delta(l)\sigma_n^2 = \sum_i\sum_j E\{a_i a_j^* X_i X_{j+l}^*\}.$$

The right-hand expected value can be expanded as in (9)

$$\delta(l)\sigma_n^2 = \sum_i\sum_j E\{a_i a_j^*\}E\{X_i X_{j+l}^*\} + A(l) + B(l) \tag{14}$$

where

$$A(l) = \sum_i\sum_j E\{a_i X_i\}E\{a_j^* X_{j+l}^*\}$$

$$B(l) = \sum_i\sum_j E\{a_i X_{j+l}^*\}E\{a_j^* X_i\}$$

for $l = -(M-1), \ldots, M-1$. The expected values in (14) can be estimated by averaging over the available data and coefficient estimates, e.g., $E\{a_j^* X_i\} = \frac{1}{N}\sum_{k=0}^{N-1}\hat{a}_j^*(k)X_i(k)$.

The spatial Fourier transform of (14) can be calculated for any array geometry, but if a linear uniform array is used, the transform can be parameterized in terms of just the direction of look θ. Therefore, the random coefficient beamformer for a uniform linear array steered in direction θ can be written as

$$P^{RC}(\theta) = \frac{\sigma_n^2 - \mathcal{A}(\theta) - \mathcal{B}(\theta)}{\sum_{i=1}^{M}\sum_{j=1}^{M} E\{a_i a_j^*\}e^{-j2\pi\frac{d}{\lambda}(i-j)\sin\theta}} \tag{15}$$

where d is the spacing between sensors, λ is the wavelength of the propagating signals, and $\mathcal{A}$ and $\mathcal{B}$ are the spatial Fourier transforms of $A(l)$ and $B(l)$, respectively, e.g.,

$$\mathcal{A}(\theta) = \sum_{l=-(M-1)}^{M-1} A(l)e^{-j2\pi\frac{d}{\lambda}l\sin\theta}.$$

3.3. Simulation Results

The random coefficient beamformer of (15) can be compared to a constant coefficient LP (CCLP) beamformer. A typical example of this comparison is shown in figure 2 where the predictive element m_0 is the first sensor for both beamformers. It is seen that while both have sharp peaks at $-10°$ and $30°$, the random coefficient beamformer has a lower noise floor with none of the spurious peaks that might be confused as signals. This improvement is due to the random coefficient model being able to fit the noisy array data better than the AR model. As was expected from the results of the hypothesis test, in the comparisons run by the authors the random coefficient beamformer clearly outperforms the CCLP almost every time and always performs at least as well as the CCLP.

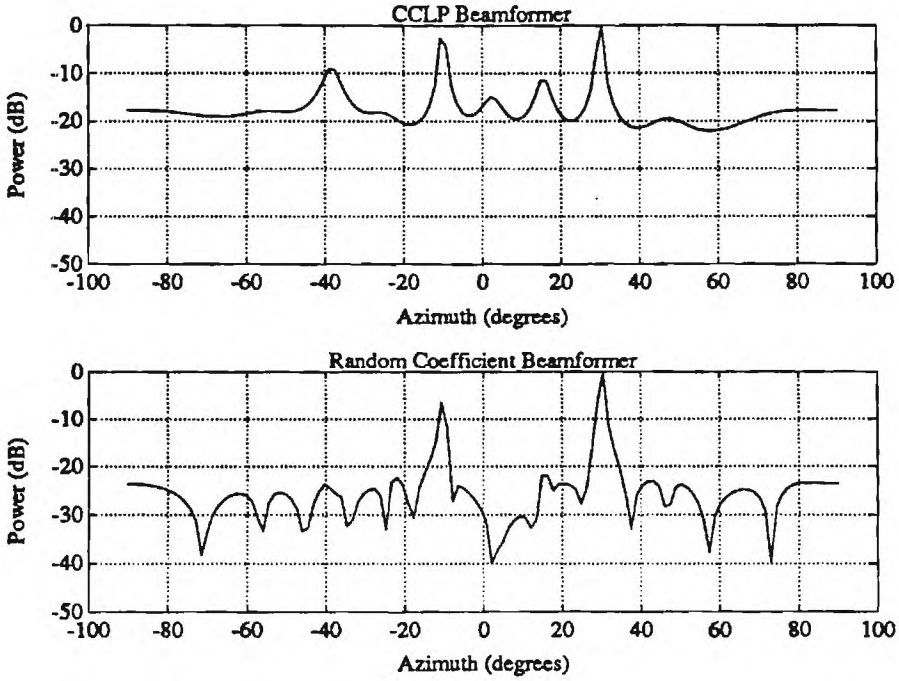

Figure 2: Comparison of the CCLP and random coefficient beamformers. The array data used was the same as that described in figure 1.

4. CONCLUSION

The random coefficient model was introduced for use in array processing. It was shown through the use of a hypothesis test that the random coefficient model will better fit typical array data than the constant coefficient autoregressive model. A beamformer using the random coefficient model was developed (15) as well as the method for estimating the random coefficients using a Kalman filter. The random coefficient beamformer was then compared to the familiar constant coefficient linear predictive beamformer and was shown to be a dramatic improvement.

REFERENCES

[1] D. H. Johnson, "The application of spectral estimation methods to bearing estimation problems," *Proc. IEEE*, vol. 70, no. 9, pp. 1018–1028, 1982.

[2] T. S. Breusch and A. R. Pagan, "A simple test for heteroscedastity and random coefficient variation," *Econometrica*, vol. 47, pp. 1287–1294, Sept. 1979.

[3] P. Newbold and T. Bos, *Stochastic Parameter Regression Models*. Beverly Hills, CA: Sage Publications, 1985.

[4] A. S. Abutaleb, "Adaptive line enhancement using a random AR model," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1211–1215, July 1990.

[5] D. F. Nicholls and B. G. Quinn, *Random Coefficient Autoregressive Models: An Introduction*. New York: Springer–Verlag, 1982.

[6] C. Hildreth and J. P. Houck, "Some estimators for a linear model with random coefficients," *J. of Amer. Stat. Ass.*, vol. 63, pp. 584–595, June 1968.

To be presented at ICASSP, San Francisco, California, March 1992.

ON ERROR FUNCTION SELECTION FOR THE ANALYSIS OF NONLINEAR TIME SERIES

Daniel F. Drake and Douglas B. Williams

School of Electrical Engineering
Georgia Institute of Technology
Atlanta, GA 30332

ABSTRACT

The extreme sensitivity of a chaotic system's steady state response to small changes in its initial conditions makes long term prediction of the evolution of such a system difficult, if not impossible. In the framework of parameter estimation, we show how this sensitivity can hinder attempts to determine model parameters that will reproduce a target chaotic time sequence. Specifically, a waveform error minimization technique based on gradient descent optimization is not well suited for estimating the parameters of a strongly chaotic system. We propose a modification of this minimization procedure that avoids some of the obstacles present when estimating the parameters of a chaotic system.

1. INTRODUCTION

Chaos—unpredictable deterministic behavior—has been observed in phenomena ranging from chemical reactions [1] to solar flares [2]. Modelling time sequences derived from such processes can provide insight into the underlying physics that drive them. Unfortunately, the intrinsic sensitivity of chaotic systems makes them difficult to model; a representation with enough freedom to correctly reproduce chaotic behavior will itself be extremely susceptible to small variations in its parameters.

The realization that long-term prediction of certain completely deterministic systems was impossible sparked interest in a new area of Dynamical Systems, an area dealing with the phenomenon of chaos. The classic description of a chaotic system usually includes the phrase "sensitive dependence on initial conditions" [3]. Sensitive, in this context, refers to the exponential rate at which initially close trajectories on the attractor diverge. This sensitivity can be quantified by the spectrum of Lyapunov exponents associated with the attractor.

Quatieri and Hofstetter [4] wished to determine the parameters and initial conditions of a nonlinear difference equation whose solution would be as close as possible to some target time sequence generated by a dynamical system. They derived a gradient descent method that minimized the waveform error between the solution of the difference equation model and the target sequence.

We will show that the waveform error surface is not well-behaved if the target sequence is generated by a chaotic system. In particular, in the neighborhood of the global minimum at least one eigenvalue of the Hessian of the waveform error increases exponentially as a function of the length of the target sequence. This unbounded growth sets the global minimum at the bottom of a deep trench, rendering gradient descent techniques impractical. Consequently, we have modified the waveform error minimization procedure by taking into account the behavior of the error surface as a function of target sequence length. This improved optimization technique provides a much wider basin of attraction for the global minimum than the original method.

This work was supported in part by the Office of Naval Research under contract N00014-91-J-4129.

2. DYNAMICAL SYSTEMS AND CHAOS

Suppose $\mathbf{h} : \mathbf{M} \times \mathbf{R}^k \rightarrow \mathbf{M}$ is a parameterized discrete-time dynamical system defined on a smooth compact manifold $\mathbf{M}$ such that

$$\mathbf{y}[n] = \mathbf{h}(\mathbf{y}[n-1], \mathbf{p}). \tag{1}$$

We assume that this system is stable, and further that the state $\mathbf{y}[n]$ converges onto an attractor $\mathbf{\Lambda} \subset \mathbf{M}$ as n tends to infinity for any initial condition $\mathbf{y}[-1]$ contained in the basin of attraction of $\mathbf{\Lambda}$.

It can be shown that the variation of the state $\mathbf{y}[n]$ with respect to initial conditions $\mathbf{y}[-1]$ is given by

$$\mathbf{D}_{\mathbf{y}[-1]}\mathbf{y}[n] = \prod_{i=-1}^{n-1} \mathbf{D}_{\mathbf{y}}\mathbf{h}(\mathbf{y}[i], \mathbf{p}). \tag{2}$$

where the operator $\mathbf{D}_{\mathbf{x}}$ applied to the vector-valued function $\mathbf{f}(\mathbf{x})$ results in a matrix with elements $(\mathbf{D}_{\mathbf{x}}\mathbf{f})_{i,j} = \partial f_i(\mathbf{x})/\partial x_j$. The Lyapunov exponents quantify the average rate at which small perturbations of the initial condition are exponentially amplified or attenuated upon iterations of the system [5]. An infinitesimal deviation $d\mathbf{y}[-1]$ will result in a deviation

$$d\mathbf{y}[n] = \prod_{i=-1}^{n-1} \mathbf{D}_{\mathbf{y}}\mathbf{h}(\mathbf{y}[i-1], \mathbf{p}) d\mathbf{y}[-1], \tag{3}$$

and for almost all $d\mathbf{y}[-1]$

$$\|d\mathbf{y}[n]\| \approx e^{\lambda(n+1)} \|d\mathbf{y}[-1]\|, \tag{4}$$

where λ is the largest Lyapunov exponent of $\mathbf{h}$ on $\mathbf{\Lambda}$. Eq (4) is equivalent to saying that $\prod_{i=-1}^{n-1} \mathbf{D}_{\mathbf{y}}\mathbf{h}(\mathbf{y}[i-1], \mathbf{p})$ has an

eigenvalue that grows on average and in absolute value as $e^{\lambda(n+1)}$. A system is, by definition, chaotic if it has least one positive Lyapunov exponent, indicating its exponential sensitivity to small variations in initial conditions.

Similarly one can show that the dependence of the state on small variations of the system's parameters is given by

$$\mathbf{D}_p\mathbf{y}[n] = \sum_{i=-1}^{n-1} \left(\prod_{j=i+1}^{n-1} \mathbf{D}_y\mathbf{h}(\mathbf{y}[j],\mathbf{p}) \right) \mathbf{D}_p\mathbf{h}(\mathbf{y}[i],\mathbf{p}). \quad (5)$$

The term $\mathbf{D}_p\mathbf{h}$ converts small deviations in the parameters into small deviations in the state which are then propagated forward by the product $\prod \mathbf{D}_y\mathbf{h}$. This coupling between parameters and state implies that a chaotic system will be extremely sensitive not only to variations in its initial conditions but to variations in its parameters as well.

3. WAVEFORM ERROR MINIMIZATION

Suppose we have a scalar time sequence $x[n] = v(\mathbf{y}[n])$ derived from a dynamical system via $v : \mathrm{M} \rightarrow \mathrm{R}$. We assume this sequence is the solution of an m^{th} order nonlinear difference equation with a known form, but depending on k unknown parameters $\mathbf{p}$. An estimate $\hat{\mathbf{p}}$ of these parameters produces the time sequence estimate

$$\hat{x}[n] = f(\hat{\mathbf{x}}[n-1], \hat{\mathbf{p}}) \text{ with } 0 \leq n < N, \quad (6)$$

where $\hat{\mathbf{x}}[n-1] = (\hat{x}[n-1], \hat{x}[n-2], \ldots, \hat{x}[n-m])^T$ is the vector of the last m values of x at time n. For simplicity we assume that the initial conditions $\mathbf{x}[-1]$ are known exactly[1]; let $\hat{\mathbf{x}}[-1] = \mathbf{x}[-1]$. We wish to find the parameters that minimize the waveform error

$$E_N = \frac{1}{N} \sum_{n=0}^{N-2} (\hat{x}[n] - x[n])^2. \quad (7)$$

Quatieri and Hofstetter use a gradient descent method to minimize the waveform error with respect to parameters. An initial estimate $\hat{\mathbf{p}}$ of the parameter values is iteratively updated

$$\hat{\mathbf{p}} \leftarrow \hat{\mathbf{p}} - \mu \left(\mathbf{D}_p E_N\right)^T \quad (8)$$

so that the error decreases at each step. The step size μ is chosen so that the error decreases at each iteration; the minimization procedure terminates when the waveform error falls below a specified threshold.

In order to better understand the behavior of the error surface, we expand E_N about the global minimum $\mathbf{p}$:

$$E_N \approx \frac{1}{N} d\mathbf{p}^T \left(\sum_{n=-1}^{N-2} \left(\mathbf{D}_p x[n]\right)^T \left(\mathbf{D}_p x[n]\right) \right) d\mathbf{p}, \quad (9)$$

where $d\mathbf{p}$ represents an infinitesimal deviation from the true parameters. A remarkable result by Takens [6] states that under the proper conditions, a scalar time sequence can be 'time delay embedded' into R^m, revealing a diffeomorphic copy of the phase space dynamics that generated the sequence. The embedding is represented by the sequence of vectors $(\mathbf{x}[n])$, where $\mathbf{x}[n] = (x[n], x[n-1], \ldots, x[n-m+1])$, with the embedding dimension m suitably chosen . The diffeomorphic relationship between the true trajectory in phase space and the reconstructed one preserves certain quantities, namely the Lyapunov exponents. Therefore the dynamical system

$$\mathbf{x}[n] = \mathbf{g}(\mathbf{x}[n-1], \mathbf{p}) = \begin{bmatrix} f(\mathbf{x}[n-1], \mathbf{p}) \\ x[n-1] \\ x[n-2] \\ \vdots \\ x[n-m+1] \end{bmatrix} \quad (10)$$

has the same Lyapunov exponents as the original dynamical system. The gradient of the scalar time sequence $(x[n])$ is simply the first row of the matrix $\mathbf{D}_p\mathbf{x}[n]$. If the dynamical system that produced the sequence is chaotic, then $\|\mathbf{D}_p x[n]\|$ will grow exponentially fast with increasing n, since it generally won't be orthogonal to the eigenvector along which the exponential expansion is taking place. Thus the Hessian of the waveform error in Eq (9), composed of a sum of outer products of the vectors $(\mathbf{D}_p x[n])$, has an increasingly large eigenvalue. As N increases the global minimum will become sandwiched between two increasingly steep walls— not ideal conditions for gradient descent optimization.

As will be seen in the next section, the waveform error seems well-behaved for *short* chaotic sequences; the exponential amplification of parameter mismatch has little time over which to markedly modify the sequence estimate. Our modification of the waveform error minimization procedure takes advantage of this phenomenon. Instead of trying to optimize our parameter estimates for the whole target sequence at once, we sequentially minimize the waveform errors $E_0, E_1, \ldots, E_N$. Once E_n sinks below some fixed threshold, we repeat the minimization process on E_{n+1}, using the last estimates of the parameters as the initial guess for the next step. Since each error surface has the global minimum in common, successive minimization of the errors forces the parameter estimates closer to their true value. We've effectively expanded the global minimum's basin of attraction to that of a length-one sequence, independent of the true sequence length.

4. EXAMPLES

A simple system capable of exhibiting chaotic behavior is the logistic equation

$$x[n] = p_1 x[n-1](1 - x[n-1]), \quad (11)$$

which is both a scalar dynamical system and a first order nonlinear difference equation. The model exhibits markedly different types of steady state behavior depending on the choice of p_1. A parameter value of $p_1 = 3.5$ results in a period-four oscillation; in contrast, a value $p_1 = 3.7$ produces chaotic behavior. Figure 1 contrasts the waveform error for both cases, for sequences of length $N = 87$ and an initial condition $x[-1] = 0.42$. While relatively flat and smooth in the periodic case, the waveform error in the chaotic case is riddled with local minima and the global

[1] The origin of the sequence can always be shifted to the right by m samples.

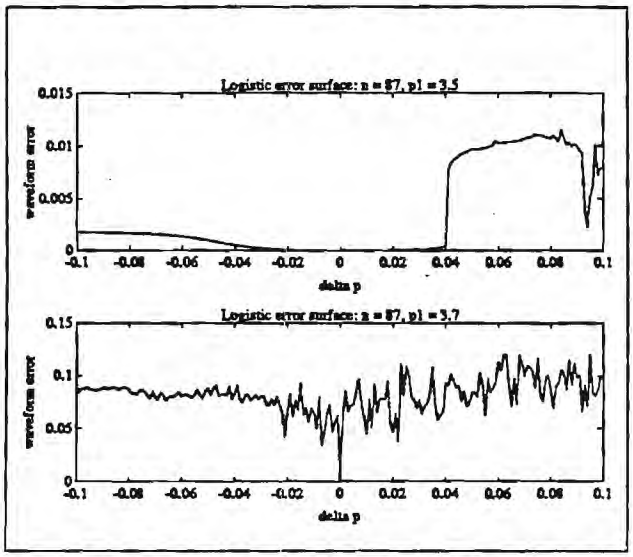

Figure 1: The waveform errors of periodic (above) and chaotic (below) target sequences of length 87, generated by the logistic equation.

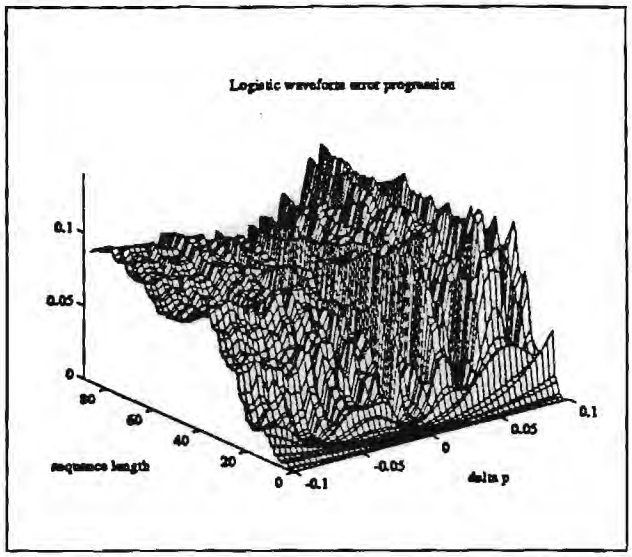

Figure 2: The waveform error as a function of target sequence length. In this case the sequence was generated by the chaotic logistic equation. The error surface is relatively smooth and shallow for short sequences, and becomes increasingly rough as more of the sequence is considered.

minimum has the very narrow basin of attraction as discussed above.

Figure 2 illustrates the behavior of the waveform error for the chaotic logistic system with respect to the sequence length. As noted previously, error surfaces for short sequence lengths are relatively smooth, giving parameter estimates with relatively large errors a greater chance of converging to the true parameter.

Figure 3 compares the performance of the original waveform error minimization technique, which tries to minimize the waveform error E_N directly, and our modified version. As expected from the general appearance of the waveform error, an initial deviation of only 5×10^{-8} in the model parameter gets trapped almost immediately in a local minimum, and the error never descends below the specified threshold of 10^{-5}. Our extension method, on the other hand, correctly identifies the true parameter after extending the target sequence to $N = 87$, even though the initial deviation from the true parameter value was 5×10^{-2}; six orders of magnitudes larger. In fact, any initial parameter value within the logistic equation's usual working range $p_0 \in [0, 4]$ will be converge to the the true parameter value.

An example of a two-parameter chaotic system is the Hénon system

$$y_1[n] = 1 - p_1 y_1^2[n-1] + y_2[n-1] \quad (12)$$

$$y_2[n] = p_2 y_1[n-1] \quad (13)$$

with parameter values $p_1 = 1.4$, $p_2 = 0.3$. If we consider the time sequence produced by the first variable ($y_1[n]$) our difference equation model has the form

$$x[n] = 1 - p_0 x^2[n-1] + p_1 x[n-2], \quad (14)$$

with initial conditions $x[-1] = y_1[-1] = 0.948586$ and $x[-2] = y_2[-1]/p_2 = 0.425317$.

Figure 4 shows that our extension method outperforms the original waveform error minimization technique. Initial parameter estimates with errors of more that 10^{-2} are reduced by six orders of magnitude. The original method, initiated with deviations of only 5×10^{-8} in both parameters, is immediately trapped in a local minimum.

Unlike the one dimensional case, for this two-parameter system our method does not produce estimates that converge to the true parameter values. Figure 5 shows the error surface E_{50} in a small neighborhood of the global minimum. As expected, the sharp gradient discussed in previous sections is in evidence. However, there also seems to be a continuous range of parameter values that generate the same waveform as those located at the global minimum. This alignment is representative of the true behavior of the Hénon system and is not an artifact of the conversion from dynamical system to difference equation. An examination of the gradient of ($y_1[n]$) with respect to the parameters shows that while the ($\mathbf{D}_p y_1[n]$) grow quickly for increasing n as expected, they also tend to align themselves along a common axis. The sum of outer products in Eq. 9 is dominated by the matrices formed from these increasingly large gradients that all point in the same direction, and consequently the Hessian appears singular. This behavior is not typical of chaotic dynamical systems in general.

5. CONCLUSION

Using concepts from the discipline of Dynamical Systems we have shown how the sensitive dependence of a chaotic system on its initial conditions can induce an analogous dependence on its parameters. Takens' embedding theorem allowed us to transplant the phase space based notion of Lyapunov exponents which quantify this sensitive depen-

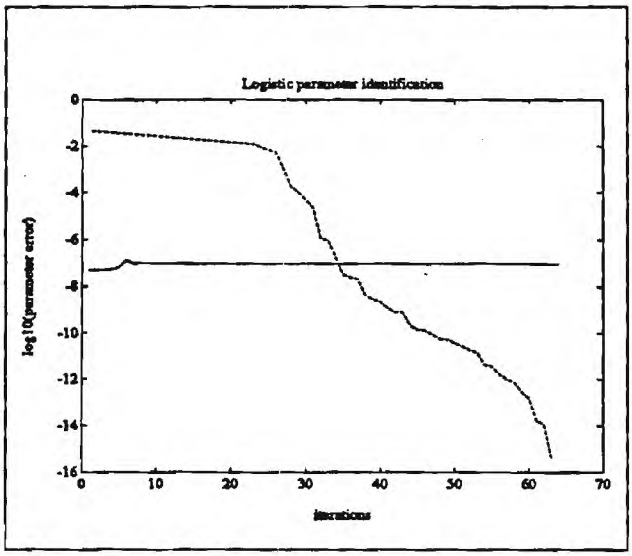

Figure 3: Comparison of waveform error minimization techniques using the chaotic logistic equation. Even good initial parameter estimates get trapped in local minima when trying to minimize the waveform error for the entire target sequence. In contrast, a much poorer initial parameter estimate converges to the true parameter value for a target sequence length of 87 when our modified minimization method is employed.

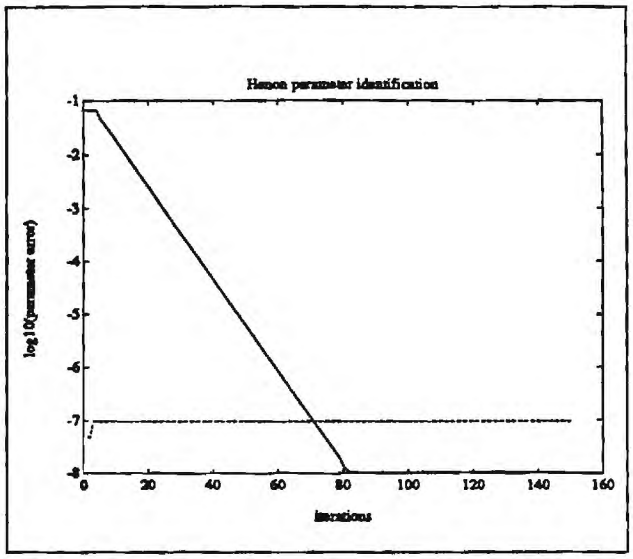

Figure 4: Parameter estimation performance for the Hénon system. Just as in the one-parameter case, the original minimization procedure procedure falls immediately into a local minimum. However, while the modified waveform error method reduces initially much larger parameter deviations better than the original, it does not converge to the true parameters, due to the singular nature of the waveform error's Hessian.

dence into a nonlinear difference equation framework. We explained how such sensitivity could produce conditions ill-suited for a proposed gradient descent minimization of the waveform error, and proposed an improved method to overcome its limitations. The improved method performed significantly better than the original when tested on sequences generated from two chaotic dynamical systems.

6. REFERENCES

[1] J.-C. Roux, R. H. Simoyi, and H. L. Swinney, "Observation of a strange attractor," *Physica D*, vol. 8, pp. 257–266, 1983.

[2] J. Kurths and H. Herzel, "An attractor in a solar time series," *Physica D*, vol. 25, pp. 165–172, 1987.

[3] J. Gleick, *Chaos*. Sphere Books Limited, 1988.

[4] T. F. Quatieri and E. M. Hofstetter, "Short-time signal representation by nonlinear difference equations," in *Proceedings of the 1990 International Conference on Acoustics, Speech, and Signal Processing (Albuquerque,NM)*, 1990.

[5] J.-P. Eckmann and D. Ruelle, "Ergodic theory of chaos and strange attractors," *Reviews of Modern Physics*, vol. 57, July 1985.

[6] F. Takens, "Detecting strange attractors in turbulence," in *Lectures Notes in Mathematics, vol. 898*, pp. 366–381, Springer, Berlin, 1981.

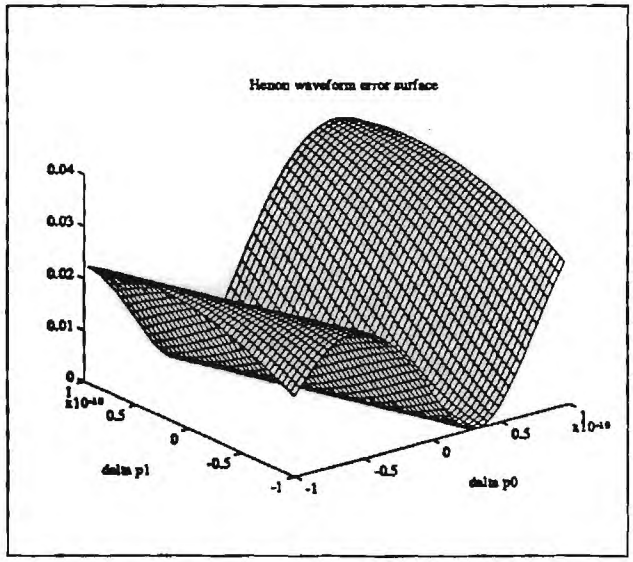

Figure 5: Waveform error surface in the neighborhood of the global minimum. The Hessian appears to be singular.

Final Technical Report
August 1, 1991 — July 31, 1992

Transform-Based Wideband Array Processing

Douglas B. Williams, Principal Investigator
Rabinder N. Madan, Scientific Officer

Contract No.: N00014–91–J–4129
R&T No.: 4148136– – –01

Over the past year, this contract has funded two projects in full and two projects in part. The two fully funded projects focus on the application of random coefficient models to wideband high-resolution direction finding and transient signal detection and estimation. The partially funded projects involve the development of a new model order determination algorithm and the analysis of nonlinear, possibly chaotic, dynamical systems. Progress in each of these areas is described below.

Random Coefficient Models: Before subspace methods such as MUSIC were developed, the highest resolution direction finding algorithms were the Minimum Variance Method and the Linear Predictive Method. The Linear Predictive Method was known to have better resolution than the Minimum Variance Method but was susceptible to bias in its direction estimates because the propagating signals and noise received by the sensors do not match the assumed autoregressive model. Consequently, the Minimum Variance Method has received much more attention and is the algorithm that most subspace methods are based on. A subspace version of the Linear Predictive Method does exist and is known to have better resolution than MUSIC, but it also has an unacceptable amount of bias.

The random coefficient model is a generalization of the autoregressive model in which the parameters are allowed to vary. We have been examining this model in the hopes of achieving resolution like that of the Linear Predictive Method but without the bias. Also, after developing the narrowband random coefficient-based algorithm, there is a natural extension to wideband direction finding. The application of random coefficient models to narrowband high resolution direction finding has been very successful and has resulted in a conference paper that was presented at ICASSP92 [1]. It has been shown that the random coefficient model is much better suited to modeling sensor array data than the autoregressive model is. A simple method for estimating the parameters of the random coefficient model has been developed and applied to simulated data. Finally, a beamformer for the random coefficient model has been developed which has significantly better performance then earlier linear predictive beamformers. Since this conference paper was written, we have developed the subspace-based version of the random coefficient beamformer. We are presently performing a detailed analysis of the bias and resolution characteristics of this beamformer and will compare them to the characteristics of MUSIC and the linear predictive methods. Though, the analysis is not yet complete, early results are very encouraging. A journal paper describing these results and the random coefficient beamformer will be submitted to the *IEEE Transactions on Signal Processing* when our analysis is complete.

We have also been examining the extension to wideband direction finding. Although there are several techniques for extending narrowband high resolution algorithms to wideband problems, none of them seem as natural as the extension of the random coefficient model for wideband applications. Each coefficient in the model simply becomes a vector spanning the range of frequencies instead of a scalar representing a single frequency. We have a method for estimating these random vector coefficients from the wideband data and are examining several techniques for extracting the direction information from the coefficients. After determining which technique is the most appropriate, we will compare it to existing wideband algorithms such as the coherent signal-subspace method.

Transient Signal Detection: Our examination of the wavelet transform and its application to the detection of transient signals with an array of sensors has led us to adaptive filter banks such as those developed by Nayebi, Barnwell, and Smith (see, for example, ICASSP 1992). These filter banks are much more general than the wavelet transform and, in fact, include the various wavelet decompositions as a subclass. The concept of an adaptive filter bank is ideal for detecting transients with unknown parameters by changing to match the received signal. Unfortunately, the adaptive filter banks that have been designed to date are for image or speech coding and are not yet suitable for transient detection. We are in the midst of specifying the constraints and filter structures to make these systems more suitable for our purposes.

After the received data has been processed by a filter bank, some additional processing is required in order to detect the presence of a signal. This processing is the topic of the appended summary [2]. This paper examines the best way to process the data efficiently after the parameters of the signal have been estimated. It also provides a way to estimate the direction-of-arrival for the signal. Because this work was done concurrently with the adaptive filter bank work, a different method (Hua and Sarkar's matrix pencil-based algorithm) was used to estimate the signal parameters. We expect it to work at least as well when the adaptive filter banks are used, especially since an adaptive filter bank requires much less computation than the SVDs required by the matrix pencil-base algorithm.

Model Order Determination: The work in this area arose by accident out of some observations made while working on another problem. However, it was supported in part by ONR as Dr. Williams's time was being supported partially by this contract. The main result of this work is the *best* method to date for determining how many signals are impinging on an array of sensors which is in some sense optimal as it minimizes the probability of making the wrong choice. This new algorithm is very similar to MDL but also has some extra terms that improve its performance significantly. Details of this method are provided in appended papers [3] and [4]. Another result of this work is that the commonly used version of MDL is incorrect and the corrected version always performs better than the previous version. Journal papers on these results are being written and will be submitted during August 1992.

Nonlinear System Identification: The majority of this work was funded by the State of Georgia, but this ONR contract was also used to fill some gaps left in the state's funding.

System identification algorithms that depend on gradient descent methods have been found to degrade significantly if the time-series or, equivalently, the system that produced the time-series is chaotic (see attached paper [5]). A careful analysis of these degradations has led to

algorithms which are much less sensitive to the potentially chaotic nature of these nonlinear systems. Analysis of these systems and their time-series continues with the eventual goal being a very general system identification algorithm to be applied to the time-series produced by nonlinear discrete-time systems.

A major hindrance to the analysis of nonlinear time-series is the inability to separate the time-series into independent, one-dimensional subsystems and thus reduce the complexity of the overall analysis. We are in the process of developing such a technique, and our progress to date is detailed in the appended summary [6]. The ability to identify and separate independent components of a nonlinear time-series should have dramatic affect on the science of nonlinear dynamical systems and could be a giant step forward in the study of nonlinear system identification. We are currently limited in the types of combinations of subsystems that can be decoupled but hope to relax these restrictions as our work progresses.

Appended Papers:

[1] B. Jost and D. B. Williams, "A beamformer based upon the random coefficient model," *Proceedings of the 1992 International Conference on Acoustics, Speech, and Signal Processing*, San Francisco, CA, vol. V, pp. 261–264, March 1992.

[2] W. T. Padgett and D. B. Williams, "Time-delay estimation for damped sinusoids incident on an array," submitted to 1993 International Conference on Acoustics, Speech, and Signal Processing, Minneapolis, MN, April 1993.

[3] D. B. Williams, "A minimum probability of error approach to designing information theoretic criteria," to appear in *Proceedings of the Fifth Digital Signal Processing Workshop*, Starved Rock State Park, IL, September 1992.

[4] D. B. Williams, "Comparison of AIC and MDL to the minimum probability of error criterion," to appear in *Proceedings of the Sixth IEEE Signal Processing Workshop on Statistical Signal and Array Processing*, Victoria, British Columbia, October 1992.

[5] D. F. Drake and D. B. Williams, "On error function selection for the analysis of nonlinear time series," *Proceedings of the 1992 International Conference on Acoustics, Speech, and Signal Processing*, San Francisco, CA, vol. V, pp. 329–332, March 1992.

[6] D. F. Drake and D. B. Williams, "Recovering dynamically independent coordinates for time-delay-embedded product systems," submitted to 1993 International Conference on Acoustics, Speech, and Signal Processing, Minneapolis, MN, April 1993.

To be presented at ICASSP, San Francisco, California, March 1992.

A BEAMFORMER BASED UPON THE RANDOM COEFFICIENT MODEL *

Bruce Jost *Douglas B. Williams*
Georgia Institute of Technology, Atlanta, GA, 30332-0250 USA

ABSTRACT

This paper applies the random coefficient model to array processing, specifically in the design of a beamformer for direction finding. This model is similar to the autoregressive (AR) model, except the coefficients are allowed to change with time instead of remaining constant; thus allowing the beamformer to better model any additive noise or signal correlations in the observations. Through the use of a binary hypothesis test, it is shown that random coefficient models better fit typical array data than do AR models. A Kalman filter is presented that has the array observations as inputs and the parameters of the random coefficient model as outputs. A new beamformer based on the random coefficient model is derived that is similar to the constant coefficient linear predictive (CCLP) beamformer. The two beamformers are compared and it is shown that the random coefficient beamformer outperforms the CCLP beamformer.

1. INTRODUCTION

The linear model is probably the most popular model in use in engineering and science because it is simple and yet powerful. Many times, computational advantages due to the simplicity of the linear model far outweigh any performance gain achieved by more complicated models. The commonly used autoregressive (AR) model is one such linear model. In array processing the AR model leads to linear predictive (LP) beamforming [1] where the output of a selected sensor, say the m_0th, is estimated as a weighted linear combination of the other sensor outputs. Assuming narrowband signals and letting $X_m(k), k = 0, \ldots, N-1$, be snapshots of the Fourier transform of the mth sensor's output at the signals' common center frequency, the LP model for $X_{m_0}(k)$ is

$$X_{m_0}(k) = -\sum_{m \neq m_0} a_m X_m(k), \qquad (1)$$

where $\{a_m\}$ is a set of complex-valued weights to be found.

When compared to other beamformers with roughly equivalent computational complexity, such as the minimum variance distortionless receiver, the LP method possesses superior resolution properties [1]. However, the performance of the LP beamformer is highly dependent on both the signals' directions of propagation and the selection of m_0. Also, in many cases the direction-of-arrival estimation bias is quite high. The primary causes of this estimation bias are additive noise in the observations and correlations between propagating signals. Additive noise and signal correlations do not obey an AR relationship across the array; so the model in (1) is no longer accurate.

In order to keep the simplicity of the linear model of (1) and yet more accurately match the array data, the random coefficient model is proposed. This model is identical to the AR model except the weights a_m are random instead of constant and (1) becomes

$$X_{m_0}(k) = -\sum_{m \neq m_0} a_m(k) X_m(k). \qquad (2)$$

The coefficients can be expressed as

$$a_m(k) = \beta_m + \nu_m(k), \qquad (3)$$

where β_m is the mean of the coefficients for the mth sensor and the $\nu_m(k)$ are zero-mean independent identically distributed random variables. The random coefficient model of (2) keeps the simple linear form of the AR model and, hence, its computational advantages. Since the constant coefficients of an AR model can be expressed in terms of (3) with the $\nu_m(k)$ equal to zero, the random coefficient model can actually be thought of as a generalization of the AR model. Therefore, the random coefficient model will fit the array data at least as well as the AR model and will simplify to the AR model when it is the better choice.

2. TESTING OF THE MODELS

Through the use of a binary hypothesis test developed by Breusch and Pagan [2], it is possible to test which model, AR or random coefficient, will better fit typical array data. The test indicates that if the residuals from an AR model have a constant variance, an AR model is better suited; otherwise, a random coefficient model should be considered. The test is performed by generating a Lagrange multiplier (LM) statistic from generated array data. If the null hypothesis of an AR model is true, then the LM statistics will be asymptotically distributed as chi-squared with M degrees of freedom, where M is the number of sensors in the array. The method used for this paper was to generate

*THIS WORK WAS SUPPORTED BY THE OFFICE OF NAVAL RESEARCH UNDER CONTRACT NUMBER N00014-91-J-4129

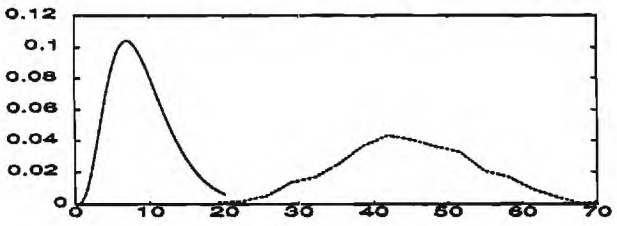

Figure 1: Results of the hypothesis test. The ten degree of freedom chi-squared distribution is the solid line while the normalized histogram of the LM statistics is the broken line. The array data was from a uniform linear array of 10 sensors with spacing $\frac{\lambda}{2}$. There are two signals present at $-10°$ and $30°$, the noise is additive white Gaussian and the SNR is 6 dB. There are 100 independent samples being used ($N = 100$).

array data for a given number of signals at their respective angles of arrival, and a given number of sensors, time samples, and signal-to-noise ratio. The LM statistic given in [2] was calculated for a large number of sets of data and the normalized histogram of these LM statistics was calculated and plotted against a chi-squared density function with M degrees of freedom to see how closely the two matched. Figure 1 shows an example of this test in which the normalized histogram of the LM statistics clearly does not match the chi-squared density plot. In fact, it was found that the AR hypothesis is rejected for even the simplest cases if any additive noise or signal correlation is present. Therefore, it would seem that the random coefficient model is better suited for array processing applications.

3. RANDOM COEFFICIENT BEAMFORMER

As the random coefficient model has been shown to fit the data more accurately than the AR model, a beamforming algorithm based on this model has been developed. The remainder of the paper will present the algorithm for estimating the random coefficients and then the derivation of the random coefficient beamformer.

3.1. Estimating the Coefficients

The first step is to estimate the random coefficients. This is performed with the Kalman filter outlined in [3] and [4]. The following set of vectors is introduced to ease the computations:

$$\begin{aligned}
\mathbf{a}'(k) &= [a_0(k)\cdots a_{m_0-1}(k)\; a_{m_0+1}(k)\cdots a_{M-1}(k)] \\
\mathbf{b}' &= [\beta_0\cdots\beta_{m_0-1}\;\beta_{m_0+1}\cdots\beta_{M-1}] \\
\mathbf{v}'(k) &= [\nu_0(k)\cdots\nu_{m_0-1}(k)\;\nu_{m_0+1}(k)\cdots\nu_{M-1}(k)] \\
\mathbf{z}'(k) &= [X_0(k)\cdots X_{m_0-1}(k)\;X_{m_0+1}(k)\cdots X_{M-1}(k)]
\end{aligned}$$

where M is the number of sensors and ($'$) denotes Hermitian transpose. An error covariance matrix is defined as

$$\Sigma = [\sigma_{ij}^2],\; i,j = 0,\ldots,m_0-1,m_0+1,\ldots,M-1$$

where $E\{\nu_i(k)\nu_j^*(l)\} = \delta(k-l)\sigma_{ij}^2$. From [3] and [4] the estimates of $\mathbf{a}(k)$ are calculated from the Kalman filter equations

$$\begin{aligned}
\hat{\mathbf{a}}(k/k-1) &= \mathbf{b} \\
\hat{\Sigma}(k/k-1) &= \Sigma \\
h(k) &= \mathbf{z}'(k)\hat{\Sigma}(k/k-1)\mathbf{z}(k)+\sigma_n^2 \\
\hat{\mathbf{a}}(k/k) &= \mathbf{b}+\hat{\Sigma}(k/k-1)\mathbf{z}(k)h^{-1}(k) \\
&\quad\cdot[X_{m_0}+\mathbf{z}'(k)\mathbf{a}(k/k-1)] \\
\hat{\Sigma}(k/k) &= \hat{\Sigma}(k/k-1)- \\
&\quad\hat{\Sigma}(k/k-1)\mathbf{z}(k)h^{-1}(k)\mathbf{z}'(k)\hat{\Sigma}(k/k-1) \qquad (4)
\end{aligned}$$

where $\hat{\mathbf{a}}(k/k)$ is the estimate of $\mathbf{a}(k)$ at time k given all the observations up to k, $\hat{\Sigma}(k/k)$ is the error covariance matrix estimate of $\mathbf{a}(k)$ at time k given all the observations up to k, and σ_n^2 is the noise variance estimate.

To implement the Kalman filter, initial estimates of $\mathbf{b}$, Σ, and σ_n^2 must be calculated. The authors will present a method that calculates the least-squares estimates of these terms following [4]. Introducing a noise term into (2) yields

$$X_{m_0}(k) = -\sum_{m\neq m_0} a_m(k)X_m(k)+n_{m_0}(k) \qquad (5)$$

where n_{m_0} is white Gaussian noise and is uncorrelated with the observations $X_m(k)$. The noise associated with an arbitrary sensor, say the mth, has the property $E\{n_m(j)n_m^*(k)\} = \delta(j-k)\sigma_n^2$. Using (3), equation (5) can be written as

$$\begin{aligned}
X_{m_0}(k) &= -\sum_{m\neq m_0}\beta_m X_m(k)-\sum_{m\neq m_0}\nu_m(k)X_m(k) \\
&\qquad +n_{m_0}(k) \\
&= -\sum_{m\neq m_0}\beta_m X_m(k)+u(k) \qquad (6)
\end{aligned}$$

where

$$u(k) = n_{m_0}(k)-\sum_{m\neq m_0}\nu_m(k)X_m(k). \qquad (7)$$

Defining

$$\begin{aligned}
\mathbf{u}' &= [u(0)\cdots u(N-1)] \\
\mathbf{y}' &= [X_{m_0}(0)\cdots X_{m_0}(N-1)] \\
\mathbf{Z}' &= [\mathbf{z}(0)\cdots\mathbf{z}(N-1)],
\end{aligned}$$

equation (6) can be written as

$$\mathbf{y} = -\mathbf{Z}\mathbf{b}+\mathbf{u}.$$

Using ordinary least squares, the estimate of $\mathbf{b}$ is

$$\hat{\mathbf{b}} = -(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}. \qquad (8)$$

Taking the expected value of the magnitude squared of (7) yields

$$E\{|u(k)|^2\} = E\{|\,n_{m_0}(k)-\sum_{m\neq m_0}\nu_m(k)X_m(k)\,|^2\}$$

which reduces to

$$E\{|u(k)|^2\} = \sigma_n^2+\sum_{i\neq m_0}\sum_{j\neq m_0}E\{\nu_i(k)\nu_j^*(k)X_i(k)X_j^*(k)\} \qquad (9)$$

where the other terms are zero because the noise is uncorrelated with the signals and coefficients. Assuming the $X_i(k)$ and $a_i(k)$ are zero mean and Gaussian, (9) can be expressed as

$$E\{|u(k)|^2\} = \sigma_n^2 + \sum_{i\neq m_0}\sum_{j\neq m_0}[E\{\nu_i(k)\nu_j^*(k)\}E\{X_i(k)X_j^*(k)\} + E\{\nu_i(k)X_i(k)\}E\{\nu_j^*(k)X_j^*(k)\} + E\{\nu_i(k)X_j^*(k)\}E\{\nu_j^*(k)X_i^*(k)\}]$$

which reduces to

$$E\{|\,u(k)\,|^2\} = \sigma_n^2 + \sum_{i\neq m_0}\sum_{j\neq m_0} E\{\nu_i(k)\nu_j^*(k)\}E\{X_i(k)X_j^*(k)\}. \quad (10)$$

Since

$$|\,u(k+l)\,|^2 = E\{|\,u(k+l)\,|^2\} + \xi_l \quad (11)$$

where the ξ_l is a zero mean error with unknown variance, (10) can be used in (11) as

$$|u(k+l)|^2 = \sigma_n^2 + \sum_{i\neq m_0}\sum_{j\neq m_0} E\{\nu_i(k+l)\nu_j^*(k+l)\}E\{X_i(k+l)X_j^*(k+l)\} + \xi_l$$

or in vector notation as

$$\mathbf{e} = \mathbf{X}\mathbf{s} + \mathbf{r}$$

where $\mathbf{e}' = [|u(0)|^2 \cdots |u(N-1)|^2], \mathbf{r}' = [\xi_0 \cdots \xi_{N-1}], \mathbf{s}' = [\sigma_n^2\ \sigma_{1,1}^2\ \sigma_{1,2}^2 \cdots \sigma_{1,M-1}^2\ \sigma_{2,1}^2 \cdots \sigma_{M-1,M-1}^2]$, and

$$\mathbf{X} = \begin{bmatrix} 1 & X_1(0)X_1^*(0) & X_1(0)X_2^*(0) & \cdots & X_1(0)X_{M-1}^*(0) & X_2(0)X_1^*(0) & \cdots & X_{M-1}(0)X_{M-1}^*(0) \\ \vdots & \vdots & & & & & & \vdots \\ 1 & X_1(N-1)X_1^*(N-1) & & \cdots & & & & X_{M-1}(N-1)X_{M-1}^*(N-1) \end{bmatrix}$$

Using ordinary least squares, the estimate of $\mathbf{s}$ is

$$\hat{\mathbf{s}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}, \quad (12)$$

from which the estimates of σ_n^2 and Σ are extracted.

3.2. The Beamformer

Once the random coefficients have been estimated, they can be used in a beamformer to calculate the direction(s) of the measured signal(s). Following the derivation of the constant coefficient LP beamformer, (5) can be rewritten as

$$n_{m_0}(k) = X_{m_0}(k) + \sum_{m\neq m_0} a_m(k)X_m(k) = \sum_{m=0}^{M-1} a_m(k)X_m(k) \quad (13)$$

if $a_{m_0}(k) = 1$ for $k = 0, \ldots, N-1$. Multiplying (13) with a similar expression for $n_{m_0+l}^*(k)$ and taking the expected value gives (dropping the time index for brevity)

$$E\{n_{m_0}n_{m_0+l}^*\} = \delta(l)\sigma_n^2 = \sum_i\sum_j E\{a_i a_j^* X_i X_{j+l}^*\}.$$

The right-hand expected value can be expanded as in (9)

$$\delta(l)\sigma_n^2 = \sum_i\sum_j E\{a_i a_j^*\}E\{X_i X_{j+l}^*\} + A(l) + B(l) \quad (14)$$

where

$$A(l) = \sum_i\sum_j E\{a_i X_i\}E\{a_j^* X_{j+l}^*\}$$

$$B(l) = \sum_i\sum_j E\{a_i X_{j+l}^*\}E\{a_j^* X_i\}$$

for $l = -(M-1), \ldots, M-1$. The expected values in (14) can be estimated by averaging over the available data and coefficient estimates, e.g., $E\{a_j^* X_i\} = \frac{1}{N}\sum_{k=0}^{N-1}\hat{a}_j^*(k)X_i(k)$.

The spatial Fourier transform of (14) can be calculated for any array geometry, but if a linear uniform array is used, the transform can be parameterized in terms of just the direction of look θ. Therefore, the random coefficient beamformer for a uniform linear array steered in direction θ can be written as

$$P^{RC}(\theta) = \frac{\sigma_n^2 - \mathcal{A}(\theta) - \mathcal{B}(\theta)}{\sum_{i=1}^{M}\sum_{j=1}^{M} E\{a_i a_j^*\}e^{-j2\pi\frac{d}{\lambda}(i-j)\sin\theta}} \quad (15)$$

where d is the spacing between sensors, λ is the wavelength of the propagating signals, and $\mathcal{A}$ and $\mathcal{B}$ are the spatial Fourier transforms of $A(l)$ and $B(l)$, respectively, e.g.,

$$\mathcal{A}(\theta) = \sum_{l=-(M-1)}^{M-1} A(l)e^{-j2\pi\frac{d}{\lambda}l\sin\theta}.$$

3.3. Simulation Results

The random coefficient beamformer of (15) can be compared to a constant coefficient LP (CCLP) beamformer. A typical example of this comparison is shown in figure 2 where the predictive element m_0 is the first sensor for both beamformers. It is seen that while both have sharp peaks at $-10°$ and $30°$, the random coefficient beamformer has a lower noise floor with none of the spurious peaks that might be confused as signals. This improvement is due to the random coefficient model being able to fit the noisy array data better than the AR model. As was expected from the results of the hypothesis test, in the comparisons run by the authors the random coefficient beamformer clearly outperforms the CCLP almost every time and always performs at least as well as the CCLP.

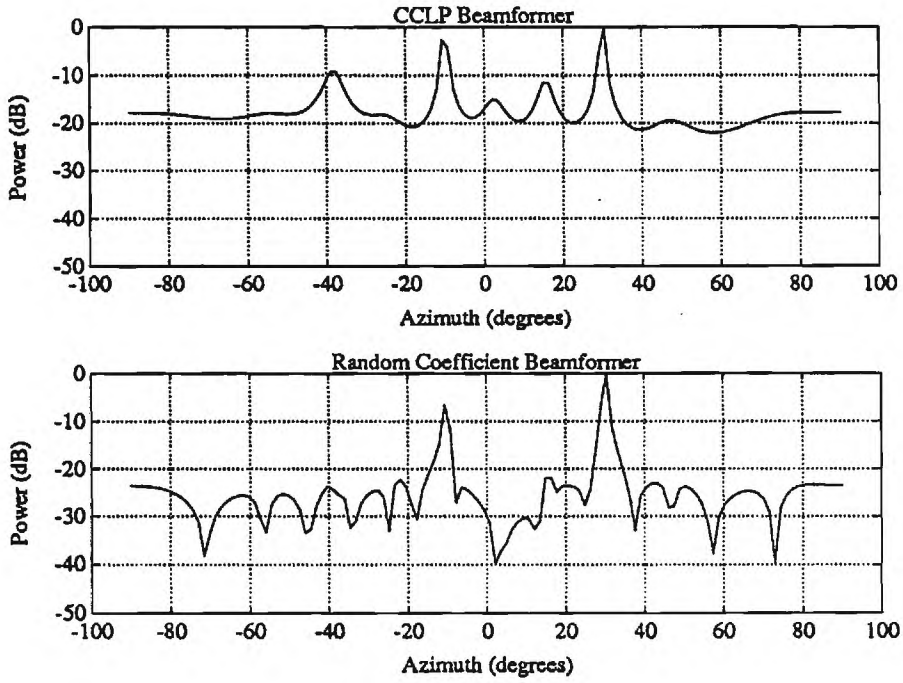

Figure 2: Comparison of the CCLP and random coefficient beamformers. The array data used was the same as that described in figure 1.

4. CONCLUSION

The random coefficient model was introduced for use in array processing. It was shown through the use of a hypothesis test that the random coefficient model will better fit typical array data than the constant coefficient autoregressive model. A beamformer using the random coefficient model was developed (15) as well as the method for estimating the random coefficients using a Kalman filter. The random coefficient beamformer was then compared to the familiar constant coefficient linear predictive beamformer and was shown to be a dramatic improvement.

REFERENCES

[1] D. H. Johnson, "The application of spectral estimation methods to bearing estimation problems," *Proc. IEEE*, vol. 70, no. 9, pp. 1018–1028, 1982.

[2] T. S. Breusch and A. R. Pagan, "A simple test for heteroscedastity and random coefficient variation," *Econometrica*, vol. 47, pp. 1287–1294, Sept. 1979.

[3] P. Newbold and T. Bos, *Stochastic Parameter Regression Models*. Beverly Hills, CA: Sage Publications, 1985.

[4] A. S. Abutaleb, "Adaptive line enhancement using a random AR model," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1211–1215, July 1990.

[5] D. F. Nicholls and B. G. Quinn, *Random Coefficient Autoregressive Models: An Introduction*. New York: Springer–Verlag, 1982.

[6] C. Hildreth and J. P. Houck, "Some estimators for a linear model with random coefficients," *J. of Amer. Stat. Ass.*, vol. 63, pp. 584–595, June 1968.

Submitted to:
ICASSP - 93
Minneapolis, MN, April 1993

Time-Delay Estimation for Damped Sinusoids Incident on an Array

Wayne T. Padgett and Douglas B. Williams
Georgia Institute of Technology
Digital Signal Processing Lab

Summary

The problem we are addressing is the estimation of intersensor time-delay for a real decaying sinusoid with unknown parameters propagating across an array of sensors. This problem has applications in the localization of signal sources by a sensor array such as a sonar or radar array. The unknown parameter assumption is designed to simulate the existence of sources of transients which are of unpredictable origin, such as the active echo-location system of another vessel, or the ringing transient generated by excitation of some natural resonant system. The location of the source of such a transient can be critical strategic information in a military situation. Previous solutions to the problem have fallen into two basic groups, those which assume a known signal spectral density, and those which assume a parametric signal model and then try to estimate the parameters.

The problem of estimating the time-delay of an unknown signal incident on an array of sensors is difficult due to the lack of information. The lack of information can be made up for by either assuming or estimating some parameters of the signal. In the first group, several good algorithms have been proposed to approach this problem assuming the signal spectral density is known. In this class, the generalized cross correlator has been shown to be the optimal solution to the two sensor problem with known spectral density [1]. In the second group, a particular form of signal is assumed, and the parameters of this form are then estimated. Most algorithms are based on the optimal maximum likelihood (ML) estimate. A slightly different approach is applied to complex signals by Hua and Sarkar with the matrix pencil based SDMP algorithm [2]. The SDMP algorithm does not depend on the ML estimate, but instead uses the fact that uniform sequential samples of a complex exponential are each powers of the same root factor or signal pole. The SDMP algorithm uses the matrix pencil to isolate and thereby estimate the signal poles. The proposed algorithm is a hybrid which uses part of the SDMP method adapted to real signals and also uses the generalized cross correlator. The cross correlator's estimate for each adjacent sensor pair is calculated and the result is averaged to determine the hybrid algorithm's final time-delay estimate. The new algorithm has several advantages including reduced computational complexity and better near and below threshold behavior than the algorithms to which it is compared. The threshold is the SNR value below which the algorithms begin to seriously degrade.

The obvious and optimal approach to dealing with a signal with unknown parameters is the ML method. This normally involves finding the ML estimate of the signal parameters by some iterative method or numerical search. Numerical searches in particular are both expensive and unreliable due to the many possible local maxima. The numerical search is also complicated in the array problem, since the time-delay variable is one of the unknown parameters, and the multiple sensors add to the size of the problem. The difficulty lies in finding a way to use all of the information available without being overwhelmed by the computation. Hua and Sarkar's SDMP algorithm can be used to estimate the signal poles in an array and with additional computation, to estimate the time delays themselves. This method takes into account all the available information and avoids the numerical search of many ML algorithms. Unfortunately, the SDMP algorithm for time-delay estimation still has the disadvantages of requiring three singular value decompositions and three generalized eigenvalue decompositions, and poor behavior below threshold. Simulations have shown that while the pole estimates are good, the time-delay estimates can have large deviations from the true value, leading to a large mean square error.

Since the matrix pencil method's first step is to estimate the poles of the time domain signal using all the array data, and we are in need of a good signal parameter estimate, we propose using the matrix pencil's

signal estimate to form the frequency window of a matched correlator. This scheme reduces the matrix pencil computations by a third, and adds a set of generalized correlators whose time-delay estimates are averaged. The window applied to the correlators is Eckhart's window which is optimum for low SNR. The spectral density assumed for computation of the window is that of the signal estimated by the first stage of the SDMP algorithm. The goal of this procedure is to use the accurate signal estimate to improve the performance of a correlator for unknown signal to the point where its performance is comparable to that of the perfectly matched correlator.

Simulations have been performed to demonstrate the improved behavior of the hybrid algorithm. Figure 1 shows a comparison of simulation results for three algorithms. The solid line shows the generalized correlator with averaged estimates and assuming the signal is known in advance. This gives a "best performance" comparison. The dotted line shows the performance of the generalized correlator with no windowing, i.e. unknown signal, to give a "worst performance" comparison. The dashed line shows the performance of the proposed hybrid algorithm. The threshold effect can be observed below 10 or 15 dB SNR. Although the behavior is erratic below threshold, it is still in the same range as the compared algorithms. Near threshold, however, the proposed algorithm begins to perform nearly as well as the known signal algorithm. The SDMP time-delay algorithm was also simulated but is not shown since its error values were too large for the scale of the plot.

In the paper and at the conference, the authors will present a detailed description and analysis of the proposed algorithm, as well highlights of any further work. The algorithm as described here is shown to be effective, but it also has the potential for generalization to the multi-signal case which would allow operation of the sensor array in a much more complex environment. We believe that the proposed algorithm in its present form represents an improvement of the state of the art in localization of unknown transients, both with respect to performance and computational load of competing algorithms.

Other Submissions

Daniel F. Drake and Douglas B. Williams, "Recovering Dynamically Independent Coordinates for Time-delay-embedded Product Systems"
S. J. Hwang and Douglas B. Williams, "A Total Least Squares Approach for Array Processing with Unknown Sensor Positions"

References

[1] G. C. Carter, "Coherence and time delay estimation," *Proceedings of the IEEE*, vol. 75, pp. 236–255, February 1987.

[2] Y. Hua and T. K. Sarkar, "Estimation of spatial and temporal parameters of multiple transient signals," in *Proc. ICASSP-90*, pp. 2411–2414, 1990.

[3] Y. Hua and T. K. Sarkar, "Generalized pencil-of-function method for extracting poles of an EM system from its transient response," *IEEE Transactions on Antennas and Propagation*, vol. 37, pp. 229–234, February 1989.

[4] J. Hassab, *Underwater Signal and Data Processing.* Boca Raton, Florida: CRC Press, 1989.

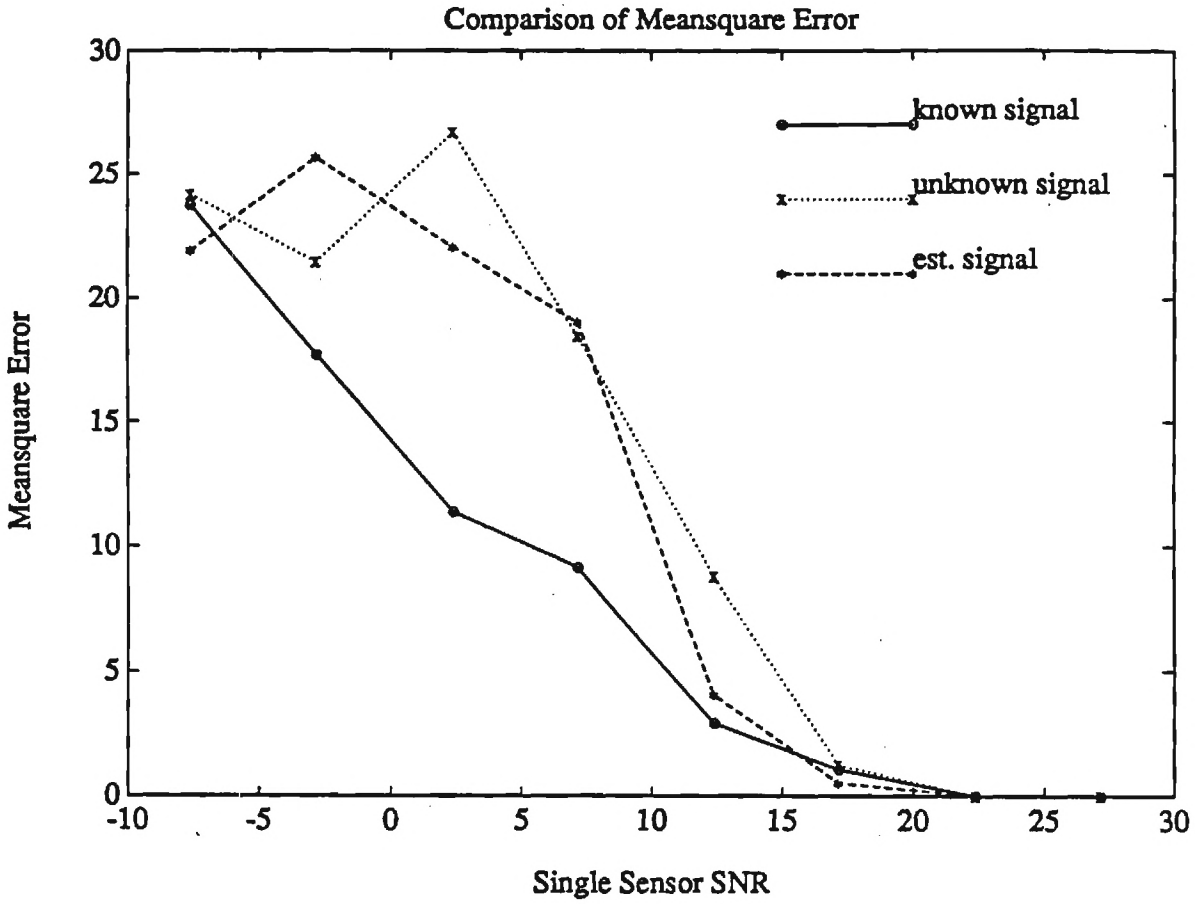

Figure 1: The simulation was carried out with the assumption of a single signal present. The simulated array has nine sensors, and the signal was the output of the system $y[k] = Re(0.7 + 0.6403i)^k$ with blocks of 48 samples collected at each sensor and an inter-sensor delay of 2. Uncorrelated additive white Gaussian noise was applied to each sensor. A total of 100 simulations were run for each SNR plotted in above, with each algorithm processing exactly the same sensor data.

To appear in:
Proceedings of the Fifth Digital Signal Processing Workshop
Starved Rock State Park, IL, September 1992

A Minimum Probability of Error Approach to Designing Information Theoretic Criteria

Douglas B. Williams
Georgia Institute of Technology
School of Electrical Engineering
Digital Signal Processing Laboratory
Atlanta, GA 30332–0250
Phone: (404) 853–9832, E-mail: dbw@eedsp.gatech.edu

A large variety of model order determination problems are solved by testing the eigenvalues of a sample covariance matrix to estimate how many of the smallest eigenvalues of the true covariance matrix are equal. Applications range from estimating the number of signals impinging on an array of sensors to determining the embedding dimension of a chaotic signal to finding the most significant data for pattern recognition. The two most popular methods for approaching these problems are AIC and MDL [2]. AIC has the advantage of good performance for those 'difficult' problems when the larger eigenvalues are not much bigger than the smallest eigenvalues but is not consistent and tends to overestimate the model order for the easier cases. MDL, on the other hand, has extremely reliable performance for most cases but falls short of AIC's performance for the difficult cases. Using the theory of multiple hypothesis tests, we derive a test that is similar to AIC and MDL and is implemented in *exactly* the same manner, but is designed to minimize the probability of choosing the wrong model order. Hence, we call this test the minimum probability of error criterion.

To arrive at our statistic, we start with the joint probability density function (*pdf*) of the eigenvalues of the $M \times M$ sample covariance when the $M-p$ smallest eigenvalues are known to be equal. We will denote this pdf by $f_p(l_1,\ldots,l_M|\lambda_1 \geq \cdots \geq \lambda_{p+1} = \cdots = \lambda_M)$ where the l_i denote the eigenvalues of the sample matrix and the λ_i are the eigenvalues of the true covariance matrix. The asymptotic expression for $f_p(\cdot)$ is given by Muirhead [1] for the real data case and by Wong *et al.* [3] for the complex data case. We then form M likelihood ratios by dividing each joint pdf by $f_{M-1}(\cdot)$ to form

$$\Lambda(p) = \frac{f_p(l_1,\ldots,l_M|\lambda_1 \geq \cdots \geq \lambda_{p+1} = \cdots = \lambda_M)}{f_{M-1}(l_1,\ldots,l_M|\lambda_1 \geq \cdots \geq \lambda_M)}, \quad p = 0, \ldots, M-1.$$

Assuming that each value of p is equally likely, then the value of p that maximizes $\Lambda(p)$ is the optimum choice in that it minimizes the probability of choosing the incorrect p. Because $\Lambda(p)$ in this form requires knowledge of the unknown parameters λ_i, we must use a generalized likelihood ratio test and independently substitute the maximum likelihood estimates of the λ_i (see [2] for these expressions) into both $f_p(\cdot)$, for which we assume $M-p$ equal λ_is, and $f_{M-1}(\cdot)$, for which we assume no equal λ_is, to get our new statistics $\Lambda(p)$. After much simplification including dropping terms that are common to $\Lambda(p)$ for every allowable value of p and then taking the natural logarithm of each $\Lambda(p)$, we get the statistics

$$\begin{aligned}\Lambda(p) &= \frac{-n+p}{2}\ln\left[\frac{\left(\frac{1}{M-p}\sum_{i=p+1}^{M} l_i\right)^{(M-p)/2}}{\prod_{i=p+1}^{M} l_i}\right] - \frac{p}{4}(2M-p-1)\ln\frac{n}{2} + \\ &\ln\left[\pi^{-p(p+1)/4}\Gamma_p(M/2)\right] + \sum_{i=1}^{p}\sum_{j=p+1}^{M}\frac{1}{2}\ln\left[\frac{l_i-l_j}{l_i-\widehat{\sigma}^2}\right] - \sum_{i=p+1}^{M}\sum_{j=i+1}^{M}\frac{1}{2}\ln\left[\frac{(l_i l_j)^{1/2}}{l_i-l_j}\right]\end{aligned}$$

This work was supported by an NSF Research Initiation Award under contract number MIP–9010790 and by the Office of Naval Research under contract number N00014–91–J–4129.

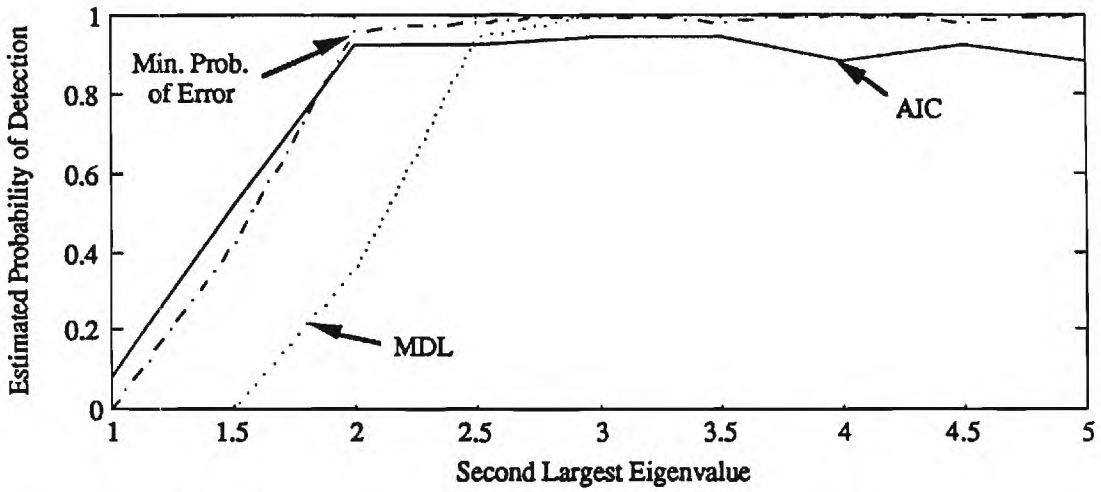

Figure 1: A comparison of the minimum probability of error criterion to AIC and MDL. Complex-valued Gaussian random vectors of length 10 were generated. The true covariance matrix for these random vectors had eight eigenvalues equal to 1, one eigenvalue equal to 5, and the tenth eigenvalue varied between 1 and 5 in increments of 0.5. For each simulation, the outer products of 100 statistically independent vectors were averaged to estimate the covariance matrix of the vectors. The estimated probability of detection is determined from the number of times out of 50 that each test correctly chose that there were 2 large eigenvalues and 8 smaller, equal eigenvalues.

---

for real data and

$$\Lambda(p) = (-n+p)\ln\left[\frac{\left(\frac{1}{M-p}\sum_{i=p+1}^{M} l_i\right)^{M-p}}{\prod_{i=p+1}^{M} l_i}\right] - \frac{p}{2}(2M-p-1)\ln[n] +$$
$$\ln\left[\frac{\pi^{-p(2M-p-1)/2}}{\widetilde{\Gamma}_{M-p}(M-p)}\right] + \sum_{i=1}^{p}\sum_{j=p+1}^{M}\ln\left[\frac{l_i - l_j}{l_i - \widehat{\sigma}^2}\right] - \sum_{i=p+1}^{M}\sum_{j=i+1}^{M} 2\ln\left[\frac{(l_i l_j)^{1/2}}{l_i - l_j}\right]$$

for complex data where n is one less than the number of samples, $\widehat{\sigma}^2 = \frac{1}{M-p}\sum_{i=p+1}^{M} l_i$, $\Gamma_N(a)$ is the multivariate gamma function [1], and $\widetilde{\Gamma}_N(a)$ is the multivariate gamma function for complex data [3].

The terms in the first line of each of these equations are almost identical to MDL. The extra terms on the following lines include both the eigenvalues being tested for equality and those not being tested. These extra terms allow the test to outperform both AIC and MDL by adapting to look like AIC for the difficult cases and like MDL otherwise. The relative performances of all three tests for complex data are illustrated above in figure 1 where it is seen that this new test statistic clearly has the best performance.

References

[1] R. J. Muirhead, "Latent Roots and Matrix Variates: A Review of Some Asymptotic Results," *The Annals of Statistics*, vol. 6, no. 1, pp. 5–33, 1978.

[2] M. Wax and T. Kailath, "Detection of Signals by Information Theoretic Criteria," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-33, pp. 387–392, Apr. 1985.

[3] K. M. Wong, Q.-T. Zhang, J. P. Reilly, and P. C. Yip, "On Information Theoretic Criteria for Determining the Number of Signals in High Resolution Array Processing," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, pp. 1959–1971, Nov. 1990.

To appear in:
Proceedings of the Sixth IEEE Signal Processing Workshop on
Statistical Signal and Array Processing
Victoria, British Columbia, October 1992

COMPARISON OF AIC AND MDL TO THE MINIMUM PROBABILITY OF ERROR CRITERION

Douglas B. Williams
Georgia Institute of Technology
School of Electrical Engineering
Digital Signal Processing Laboratory
Atlanta, GA 30332–0250

ABSTRACT

A large variety of model order determination problems involve testing the eigenvalues of a sample covariance matrix to estimate how many of the smallest eigenvalues of the true covariance matrix are equal. The two most popular methods are AIC and MDL. Using the theory of multiple hypothesis tests, we derive the *minimum probability of error criterion* that is similar to AIC and MDL and is implemented in *exactly* the same manner, but is designed to minimize the probability of choosing the wrong model order. The basic structure of this test is very similar to that of AIC and MDL except for an extra term that increases adaptability and enables this criterion to outperform both AIC and MDL.

1. INTRODUCTION

Many model order determination problems are basically tests on the eigenvalues of a sample covariance matrix to determine how many of the smallest eigenvalues of the true covariance matrix are equal. Applications range from estimating the number of signals impinging on an array of sensors to determining the embedding dimension of a chaotic signal to finding the most significant data for pattern recognition. AIC and MDL [5] are easily the most popular methods for solving these problems. For real-valued data these two techniques consist of picking the smallest of

$$AIC(p) = -(M-p)N \ln \left[\frac{\prod_{i=p+1}^{M} l_i^{1/(M-p)}}{\frac{1}{M-p}\sum_{i=p+1}^{M} l_i} \right] + p(2M+1-p)$$

or

$$MDL(p) = -\frac{1}{2}(M-p)N \ln \left[\frac{\prod_{i=p+1}^{M} l_i^{1/(M-p)}}{\frac{1}{M-p}\sum_{i=p+1}^{M} l_i} \right] + \frac{1}{4}p(2M+1-p)\ln N$$

for $p = 0, \ldots, M-1$ where the l_i are the eigenvalues in descending order of the $M \times M$ sample covariance matrix and N is the number of sample vectors. The sample covariance matrix is found by averaging the outer products of the sample vectors. For complex-valued data, simply choose the smallest of [5]

$$AIC(p) = -2(M-p)N \ln \left[\frac{\prod_{i=p+1}^{M} l_i^{1/(M-p)}}{\frac{1}{M-p}\sum_{i=p+1}^{M} l_i} \right] + 2p(2M-p)$$

or

$$MDL(p) = -(M-p)N \ln \left[\frac{\prod_{i=p+1}^{M} l_i^{1/(M-p)}}{\frac{1}{M-p}\sum_{i=p+1}^{M} l_i} \right] + \frac{1}{2}p(2M-p)\ln N$$

for $p = 0, \ldots, M-1$. AIC has the advantage of good performance for those 'difficult' problems when the larger eigenvalues are not much bigger than the smallest eigenvalues but is not consistent and tends to overestimate the model order for the easier cases. MDL, on the other hand, has extremely reliable performance for most cases but falls short of AIC's performance for the difficult cases. Using the theory of multiple hypothesis tests, we derive a test that is similar to AIC and MDL and is implemented in *exactly* the same manner, but is designed to minimize the probability of choosing the wrong model order. Hence, we call this test the minimum probability of error criterion. We show that our new criterion has significantly better performance than either AIC or MDL. Additionally, the form of our new criterion suggests some simple modifications to MDL that measurably increase its performance.

2. A MINIMUM PROBABILITY OF ERROR APPROACH

To arrive at our statistic, we start with the joint probability density function (*pdf*) of the eigenvalues of the $M \times M$ sample covariance when the $M-p$ smallest eigenvalues are known to be equal. We will denote this pdf by $f_p(l_1, \ldots, l_M | \lambda_1 \geq \cdots \geq \lambda_{p+1} = \cdots = \lambda_M)$ where the l_i denote the eigenvalues of the sample covariance matrix and the λ_i are the eigenvalues of the true covariance matrix. The asymptotic expression for $f_p(\cdot)$ is given by Muirhead [4] for the real-valued data case as

This work was supported by an NSF Research Initiation Award under contract number MIP–9010790 and by the Office of Naval Research under contract number N00014–91–J–4129.

$$f_p(l_1,\ldots,l_M|\lambda_1 \geq \cdots \geq \lambda_{p+1} = \cdots = \lambda_M) \approx$$
$$\frac{\left(\frac{n}{2}\right)^{\frac{mn}{2}-\frac{p}{4}(2M-p-1)}\pi^{\frac{M^2}{2}-\frac{p}{4}(p+1)}\Gamma_p\left(\frac{M}{2}\right)}{\Gamma_M(n/2)\Gamma_M(M/2)}\prod_{i=1}^{M}\lambda_i^{-n/2}$$
$$\prod_{i=1}^{M} l_i^{\frac{n-M-1}{2}}\exp\left\{-\frac{n}{2}\sum_{i=1}^{M}\frac{l_i}{\lambda_i}\right\}\prod_{i=p+1}^{M}\prod_{i<j}^{M}(l_i-l_j)$$
$$\prod_{i=1}^{p}\prod_{i<j}^{p}\left(\frac{(l_i-l_j)\lambda_i\lambda_j}{\lambda_i-\lambda_j}\right)^{\frac{1}{2}}\prod_{i=1}^{p}\prod_{j=p+1}^{M}\left(\frac{(l_i-l_j)\lambda_i\lambda_j}{\lambda_i-\lambda_j}\right)^{\frac{1}{2}}$$

and by Chattopadhyay [3] and Wong *et al.* [6] for the complex-valued data case as

$$f_p(l_1,\ldots,l_M|\lambda_1 \geq \cdots \geq \lambda_{p+1} = \cdots = \lambda_M) \approx$$
$$\frac{\pi^{mn-\frac{p}{2}(2M-p-1)}\pi^{M(M-1)-\frac{p}{2}(2M-p-1)}}{\widetilde{\Gamma}_M(n)\widetilde{\Gamma}_{M-p}(M-p)}\prod_{i=1}^{M}\lambda_i^{-n}$$
$$\prod_{i=1}^{M} l_i^{n-M}\exp\left\{-n\sum_{i=1}^{M}\frac{l_i}{\lambda_i}\right\}\prod_{i=p+1}^{M}\prod_{i<j}^{M}(l_i-l_j)^2$$
$$\prod_{i=1}^{p}\prod_{i<j}^{p}\left(\frac{(l_i-l_j)\lambda_i\lambda_j}{\lambda_i-\lambda_j}\right)\prod_{i=1}^{p}\prod_{j=p+1}^{M}\left(\frac{(l_i-l_j)\lambda_i\lambda_j}{\lambda_i-\lambda_j}\right)$$

where $\Gamma_N(a)$ is the multivariate gamma function [4], and $\widetilde{\Gamma}_N(a)$ is the multivariate gamma function for complex-valued data [6]. We then form M likelihood ratios by dividing each joint pdf by $f_{M-1}(\cdot)$ to form

$$\Lambda(p) = \frac{f_p(l_1,\ldots,l_M|\lambda_1 \geq \cdots \geq \lambda_{p+1} = \cdots = \lambda_M)}{f_{M-1}(l_1,\ldots,l_M|\lambda_1 \geq \cdots \geq \lambda_M)}, \quad p = 0,\ \ldots,\ M-1.$$

Assuming that each value of p is equally likely, then the value of p that maximizes $\Lambda(p)$ is the optimum choice in that it minimizes the probability of choosing the incorrect p. Because $\Lambda(p)$ in this form requires knowledge of the unknown parameters λ_i, we must use a generalized likelihood ratio test. To get our new statistics $\Lambda(p)$, we independently substitute the maximum likelihood estimates of the λ_i into both joint pdfs. For $f_p(\cdot)$ we assume $M-p$ equal λ_is and the maximum likelihood estimates are [1]

$$\widehat{\lambda}_i = \begin{cases} l_i; & i = 1,\ldots,p \\ \widehat{\sigma}^2 = \frac{1}{M-p}\sum_{i=p+1}^{M} l_i; & i = p+1,\ldots,M \end{cases}$$

For $f_{M-1}(\cdot)$ we assume no equal λ_is and the maximum likelihood estimates are simply $\widehat{\lambda}_i = l_i$. Wong *et al.* [6] have derived different values for λ_i that maximize the asymptotic density $f_p(\cdot)$ given above for the complex-valued data case. However, we have chosen not to use their expressions for several reasons: their expressions maximize the asymptotic density and not the actual density and, consequently, are not the maximum likelihood estimates; asymptotically, their expressions are equivalent to the maximum likelihood estimates; and their expressions are not in closed form but are actually a set of nonlinear equations to be solved using Newton's Method and, as such, would be very impractical for substitution into $\Lambda(p)$.

After much simplification, including dropping terms that are common to $\Lambda(p)$ for every allowable value of p and then taking the natural logarithm of each $\Lambda(p)$, we get the statistics

$$\Lambda(p) = \frac{-n+p}{2}\ln\left[\frac{\left(\frac{1}{M-p}\sum_{i=p+1}^{M} l_i\right)^{M-p}}{\prod_{i=p+1}^{M} l_i}\right] -$$
$$\frac{p}{4}(2M-p-1)\ln\frac{n}{2} + \ln\left[\pi^{-\frac{p(p+1)}{4}}\Gamma_p(M/2)\right] +$$
$$\sum_{i=1}^{p}\sum_{j=p+1}^{M}\frac{1}{2}\ln\left[\frac{l_i-l_j}{l_i-\widehat{\sigma}^2}\right] - \sum_{i=p+1}^{M}\sum_{j=i+1}^{M}\ln\left[\frac{(l_il_j)^{1/2}}{l_i-l_j}\right]$$

for real-valued data and

$$\Lambda(p) = (-n+p)\ln\left[\frac{\left(\frac{1}{M-p}\sum_{i=p+1}^{M} l_i\right)^{M-p}}{\prod_{i=p+1}^{M} l_i}\right] -$$
$$\frac{p}{2}(2M-p-1)\ln n + \ln\left[\frac{\pi^{-\frac{p(2M-p-1)}{2}}}{\widetilde{\Gamma}_{M-p}(M-p)}\right] +$$
$$\sum_{i=1}^{p}\sum_{j=p+1}^{M}\ln\left[\frac{l_i-l_j}{l_i-\widehat{\sigma}^2}\right] - \sum_{i=p+1}^{M}\sum_{j=i+1}^{M} 2\ln\left[\frac{(l_il_j)^{1/2}}{l_i-l_j}\right]$$

for complex-valued data where n is one less than the number of samples and $\widehat{\sigma}^2 = \frac{1}{M-p}\sum_{i=p+1}^{M} l_i$. The minimum probability of error criterion consists of choosing the value of p that maximizes $\Lambda(p)$ over $p = 0,\ldots,M-1$.

The terms in the first line of each of these equations are almost identical to MDL. The extra terms on the following lines include both the eigenvalues being tested for equality and those not being tested. These extra terms allow the test to outperform both AIC and MDL by adapting to look like AIC for the difficult cases and like MDL otherwise. The relative performances of all three tests are displayed in figure 1 for real-valued data and in figure 2 for complex-valued data where it is seen that this new test criterion clearly has the best performance.

The similarity between the first two terms in the minimum probability of error criterion and the statistics for MDL suggest that the performance of MDL may be improved by modifying it to be *exactly* the first terms in the minimum probability of error criterion. These adjustments yield what we call the modified MDL method which consists of minimizing

$$modMDL(p) = -\frac{1}{2}(M-p)(n-p)\ln\left[\frac{\prod_{i=p+1}^{M} l_i^{1/(M-p)}}{\frac{1}{M-p}\sum_{i=p+1}^{M} l_i}\right] + \frac{1}{4}p(2M-1-p)\ln n$$

over $p = 0,\ldots,M-1$ for the real-valued case or minimizing exactly twice this statistic over $p = 0,\ldots,M-1$ for the complex-valued case. Thus, exactly the same test may be

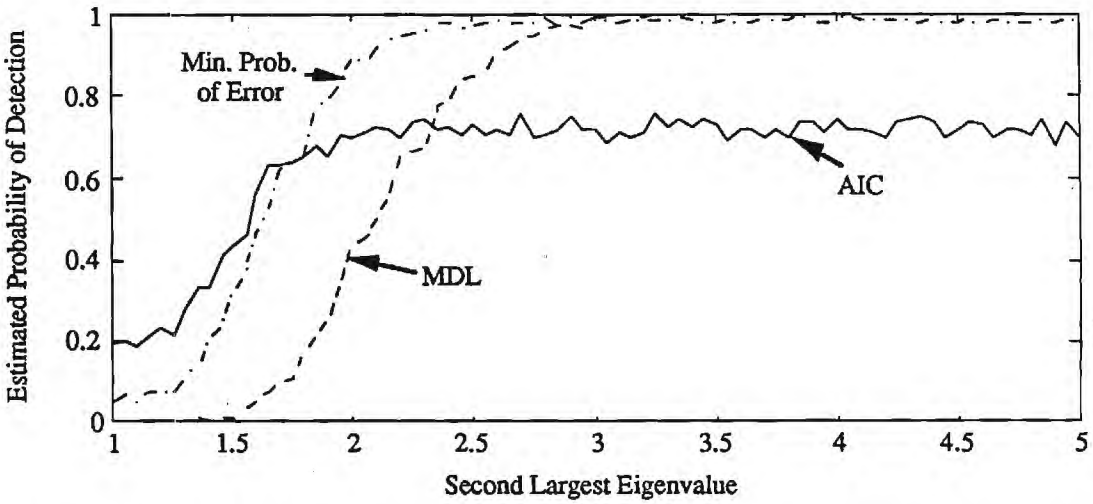

Figure 1: A comparison of the minimum probability of error criterion to AIC and MDL. Real-valued Gaussian random vectors of length 10 were generated. The covariance matrix for these random vectors had eight eigenvalues equal to 1, one eigenvalue equal to 5, and the tenth eigenvalue varied between 1 and 5 in increments of 0.05. For each simulation, the outer products of 100 statistically independent vectors were averaged to estimate the covariance matrix of the vectors. The estimated probability of detection is determined from the number of times out of 500 that each test correctly chose that there were 2 large eigenvalues and 8 smaller, equal eigenvalues.

used for both real-valued and complex-valued data. The substitution of $n-p$ for N is, interestingly enough, the same substitution proposed by Bartlett [2] for modifying the sphericity test to test subsets of the sample eigenvalues for equality. The adjustment to the second term suggests that Wax and Kailath [5] miscounted the degrees of freedom in these model order determination problems. In fact, it is possible to show that this new value of $2M-p-1$ reflects the true number of degrees of freedom. The relative performance of MDL and the modified MDL are displayed in figure 3 for complex-valued data. As can be seen, the modified MDL always performs better than MDL, though perhaps not remarkably so.

3. CONCLUSIONS

We have derived a new criterion for model order determination that has the same basic structure as AIC and MDL and is optimum in that it minimizes the probability of choosing the wrong model order. However, this new criterion is only applicable to those model order problems in which the eigenvalues of the sample covariance matrix are tested for equality and, therefore, is not as versatile as the general expressions for AIC and MDL.

The minimum probability of error criterion was demonstrated via simulations to combine the strong points of AIC and MDL. Extra terms in this new criterion allow it to perform both like AIC when AIC performs well and like MDL when MDL performs well. Because it tends to perform at least as well as either AIC or MDL, it performs better than either method individually.

Because of the strong similarity between this new criterion and MDL, we have suggested minor adjustments to MDL. These modifications were shown to result in a version of MDL that always performs at least as well and sometimes better than the previous version. Consequently, if MDL is to be implemented, we would recommend using this modified form.

REFERENCES

[1] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis.* New York: John Wiley and Sons, second ed., 1984.

[2] M. S. Bartlett, "A Note on the Multiplying Factors for Various χ^2 Approximations," *Journal of the Royal Statistical Society, Series B*, vol. 16, pp. 296–298, 1954.

[3] A. K. Chattopadhyay, "An Asymptotic Distribution Theory and Applications in Multivariate Analysis," Tech. Rep. 256, Dept. of Statistics, Purdue Univ., 1971.

[4] R. J. Muirhead, "Latent Roots and Matrix Variates: A Review of Some Asymptotic Results," *The Annals of Statistics*, vol. 6, no. 1, pp. 5–33, 1978.

[5] M. Wax and T. Kailath, "Detection of Signals by Information Theoretic Criteria," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-33, pp. 387–392, Apr. 1985.

[6] K. M. Wong, Q.-T. Zhang, J. P. Reilly, and P. C. Yip, "On Information Theoretic Criteria for Determining the Number of Signals in High Resolution Array Processing," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, pp. 1959–1971, Nov. 1990.

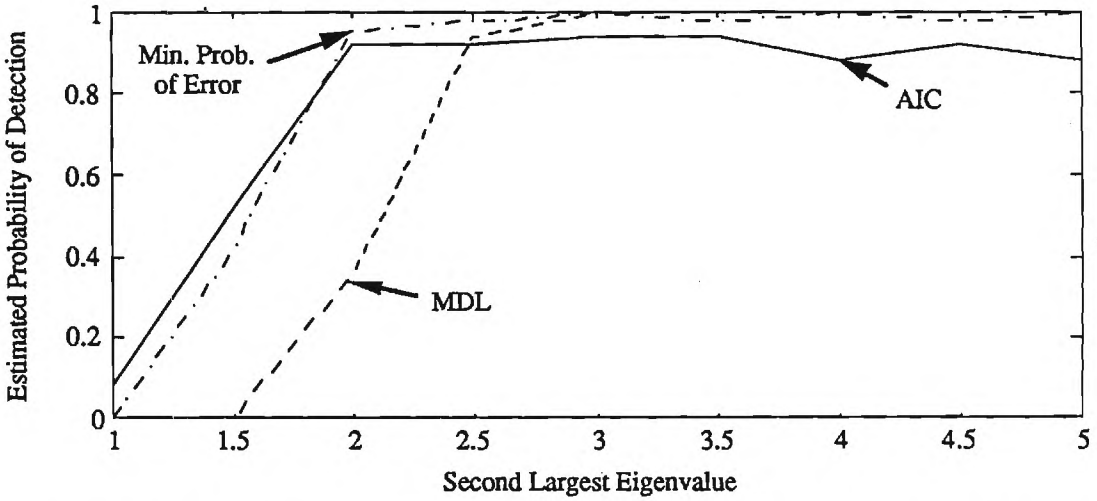

Figure 2: A comparison of the minimum probability of error criterion to AIC and MDL. Complex-valued Gaussian random vectors of length 10 were generated. The true covariance matrix for these random vectors had eight eigenvalues equal to 1, one eigenvalue equal to 5, and the tenth eigenvalue varied between 1 and 5 in increments of 0.5. For each simulation, the outer products of 100 statistically independent vectors were averaged to estimate the covariance matrix of the vectors. The estimated probability of detection is determined from the number of times out of 50 that each test correctly chose that there were 2 large eigenvalues and 8 smaller, equal eigenvalues.

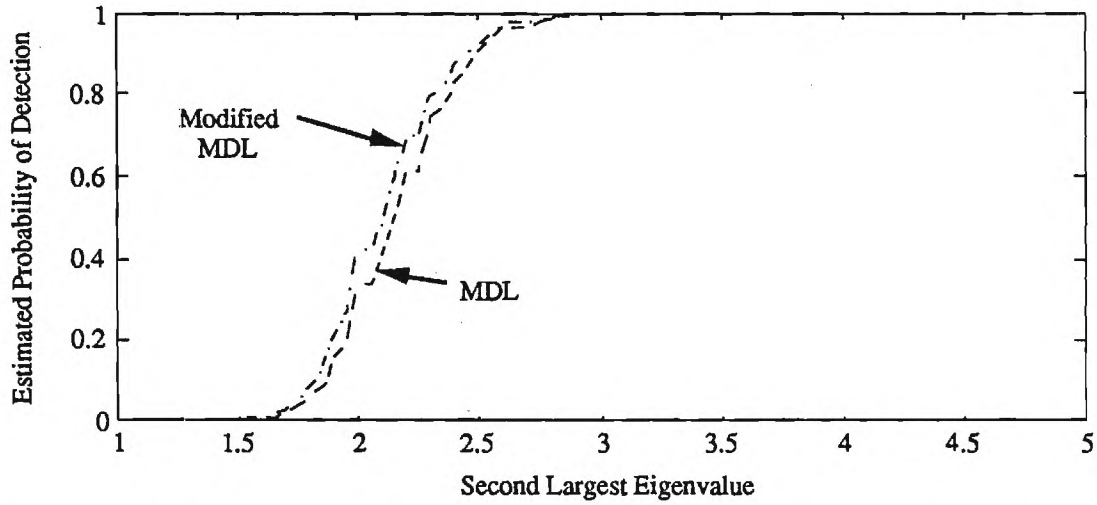

Figure 3: A comparison of the modified MDL criterion to MDL. Complex-valued Gaussian random vectors of length 10 were generated. The true covariance matrix for these random vectors had eight eigenvalues equal to 1, one eigenvalue equal to 5, and the tenth eigenvalue varied between 1 and 5 in increments of 0.05. For each simulation, the outer products of 100 statistically independent vectors were averaged to estimate the covariance matrix of the vectors. The estimated probability of detection is determined from the number of times out of 500 that each test correctly chose that there were 2 large eigenvalues and 8 smaller, equal eigenvalues.

To be presented at ICASSP, San Francisco, California, March 1992.

ON ERROR FUNCTION SELECTION FOR THE ANALYSIS OF NONLINEAR TIME SERIES

Daniel F. Drake and Douglas B. Williams

School of Electrical Engineering
Georgia Institute of Technology
Atlanta, GA 30332

ABSTRACT

The extreme sensitivity of a chaotic system's steady state response to small changes in its initial conditions makes long term prediction of the evolution of such a system difficult, if not impossible. In the framework of parameter estimation, we show how this sensitivity can hinder attempts to determine model parameters that will reproduce a target chaotic time sequence. Specifically, a waveform error minimization technique based on gradient descent optimization is not well suited for estimating the parameters of a strongly chaotic system. We propose a modification of this minimization procedure that avoids some of the obstacles present when estimating the parameters of a chaotic system.

1. INTRODUCTION

Chaos—unpredictable deterministic behavior—has been observed in phenomena ranging from chemical reactions [1] to solar flares [2]. Modelling time sequences derived from such processes can provide insight into the underlying physics that drive them. Unfortunately, the intrinsic sensitivity of chaotic systems makes them difficult to model; a representation with enough freedom to correctly reproduce chaotic behavior will itself be extremely susceptible to small variations in its parameters.

The realization that long-term prediction of certain completely deterministic systems was impossible sparked interest in a new area of Dynamical Systems, an area dealing with the phenomenon of chaos. The classic description of a chaotic system usually includes the phrase "sensitive dependence on initial conditions" [3]. Sensitive, in this context, refers to the exponential rate at which initially close trajectories on the attractor diverge. This sensitivity can be quantified by the spectrum of Lyapunov exponents associated with the attractor.

Quatieri and Hofstetter [4] wished to determine the parameters and initial conditions of a nonlinear difference equation whose solution would be as close as possible to some target time sequence generated by a dynamical system. They derived a gradient descent method that minimized the waveform error between the solution of the difference equation model and the target sequence.

We will show that the waveform error surface is not well-behaved if the target sequence is generated by a chaotic system. In particular, in the neighborhood of the global minimum at least one eigenvalue of the Hessian of the waveform error increases exponentially as a function of the length of the target sequence. This unbounded growth sets the global minimum at the bottom of a deep trench, rendering gradient descent techniques impractical. Consequently, we have modified the waveform error minimization procedure by taking into account the behavior of the error surface as a function of target sequence length. This improved optimization technique provides a much wider basin of attraction for the global minimum than the original method.

This work was supported in part by the Office of Naval Research under contract N00014-91-J-4129.

2. DYNAMICAL SYSTEMS AND CHAOS

Suppose $\mathbf{h} : \mathrm{M} \times \mathrm{R}^k \rightarrow \mathrm{M}$ is a parameterized discrete-time dynamical system defined on a smooth compact manifold M such that

$$\mathbf{y}[n] = \mathbf{h}(\mathbf{y}[n-1], \mathbf{p}). \tag{1}$$

We assume that this system is stable, and further that the state $\mathbf{y}[n]$ converges onto an attractor $\Lambda \subset \mathrm{M}$ as n tends to infinity for any initial condition $\mathbf{y}[-1]$ contained in the basin of attraction of Λ.

It can be shown that the variation of the state $\mathbf{y}[n]$ with respect to initial conditions $\mathbf{y}[-1]$ is given by

$$\mathbf{D}_{\mathbf{y}[-1]}\mathbf{y}[n] = \prod_{i=-1}^{n-1} \mathbf{D}_{\mathbf{y}}\mathbf{h}(\mathbf{y}[i], \mathbf{p}). \tag{2}$$

where the operator $\mathbf{D}_{\mathbf{x}}$ applied to the vector-valued function $\mathbf{f}(\mathbf{x})$ results in a matrix with elements $(\mathbf{D}_{\mathbf{x}}\mathbf{f})_{i,j} = \partial f_i(\mathbf{x})/\partial x_j$. The Lyapunov exponents quantify the average rate at which small perturbations of the initial condition are exponentially amplified or attenuated upon iterations of the system [5]. An infinitesimal deviation $d\mathbf{y}[-1]$ will result in a deviation

$$d\mathbf{y}[n] = \prod_{i=-1}^{n-1} \mathbf{D}_{\mathbf{y}}\mathbf{h}(\mathbf{y}[i-1], \mathbf{p}) d\mathbf{y}[-1], \tag{3}$$

and for almost all $d\mathbf{y}[-1]$

$$\|d\mathbf{y}[n]\| \approx e^{\lambda(n+1)} \|d\mathbf{y}[-1]\|, \tag{4}$$

where λ is the largest Lyapunov exponent of $\mathbf{h}$ on Λ. Eq (4) is equivalent to saying that $\prod_{i=-1}^{n-1} \mathbf{D}_{\mathbf{y}}\mathbf{h}(\mathbf{y}[i-1], \mathbf{p})$ has an

eigenvalue that grows on average and in absolute value as $e^{\lambda(n+1)}$. A system is, by definition, chaotic if it has least one positive Lyapunov exponent, indicating its exponential sensitivity to small variations in initial conditions.

Similarly one can show that the dependence of the state on small variations of the system's parameters is given by

$$\mathbf{D}_p\mathbf{y}[n] = \sum_{i=-1}^{n-1} \left(\prod_{j=i+1}^{n-1} \mathbf{D}_y\mathbf{h}(\mathbf{y}[j],\mathbf{p}) \right) \mathbf{D}_p\mathbf{h}(\mathbf{y}[i],\mathbf{p}). \quad (5)$$

The term $\mathbf{D}_p\mathbf{h}$ converts small deviations in the parameters into small deviations in the state which are then propagated forward by the product $\prod \mathbf{D}_y\mathbf{h}$. This coupling between parameters and state implies that a chaotic system will be extremely sensitive not only to variations in its initial conditions but to variations in its parameters as well.

3. WAVEFORM ERROR MINIMIZATION

Suppose we have a scalar time sequence $x[n] = v(\mathbf{y}[n])$ derived from a dynamical system via $v : \mathrm{M} \to \mathrm{R}$. We assume this sequence is the solution of an m^{th} order nonlinear difference equation with a known form, but depending on k unknown parameters $\mathbf{p}$. An estimate $\hat{\mathbf{p}}$ of these parameters produces the time sequence estimate

$$\hat{x}[n] = f(\hat{\mathbf{x}}[n-1], \hat{\mathbf{p}}) \text{ with } 0 \le n < N, \quad (6)$$

where $\hat{\mathbf{x}}[n-1] = (\hat{x}[n-1], \hat{x}[n-2], \ldots, \hat{x}[n-m])^T$ is the vector of the last m values of x at time n. For simplicity we assume that the initial conditions $\mathbf{x}[-1]$ are known exactly[1]; let $\hat{\mathbf{x}}[-1] = \mathbf{x}[-1]$. We wish to find the parameters that minimize the waveform error

$$E_N = \frac{1}{N} \sum_{n=0}^{N-2} (\hat{x}[n] - x[n])^2. \quad (7)$$

Quatieri and Hofstetter use a gradient descent method to minimize the waveform error with respect to parameters. An initial estimate $\hat{\mathbf{p}}$ of the parameter values is iteratively updated

$$\hat{\mathbf{p}} \leftarrow \hat{\mathbf{p}} - \mu \left(\mathbf{D}_p E_N\right)^T \quad (8)$$

so that the error decreases at each step. The step size μ is chosen so that the error decreases at each iteration; the minimization procedure terminates when the waveform error falls below a specified threshold.

In order to better understand the behavior of the error surface, we expand E_N about the global minimum $\mathbf{p}$:

$$E_N \approx \frac{1}{N} d\mathbf{p}^T \left(\sum_{n=-1}^{N-2} \left(\mathbf{D}_p x[n]\right)^T \left(\mathbf{D}_p x[n]\right) \right) d\mathbf{p}, \quad (9)$$

where $d\mathbf{p}$ represents an infinitesimal deviation from the true parameters. A remarkable result by Takens [6] states that under the proper conditions, a scalar time sequence can be 'time delay embedded' into R^m, revealing a diffeomorphic copy of the phase space dynamics that generated the sequence. The embedding is represented by the sequence of vectors $(\mathbf{x}[n])$, where $\mathbf{x}[n] = (x[n], x[n-1], \ldots, x[n-m+1])$, with the embedding dimension m suitably chosen . The diffeomorphic relationship between the true trajectory in phase space and the reconstructed one preserves certain quantities, namely the Lyapunov exponents. Therefore the dynamical system

$$\mathbf{x}[n] = \mathbf{g}(\mathbf{x}[n-1], \mathbf{p}) = \begin{bmatrix} f(\mathbf{x}[n-1], \mathbf{p}) \\ x[n-1] \\ x[n-2] \\ \vdots \\ x[n-m+1] \end{bmatrix} \quad (10)$$

has the same Lyapunov exponents as the original dynamical system. The gradient of the scalar time sequence $(x[n])$ is simply the first row of the matrix $\mathbf{D}_p\mathbf{x}[n]$. If the dynamical system that produced the sequence is chaotic, then $\|\mathbf{D}_p x[n]\|$ will grow exponentially fast with increasing n, since it generally won't be orthogonal to the eigenvector along which the exponential expansion is taking place. Thus the Hessian of the waveform error in Eq (9), composed of a sum of outer products of the vectors $(\mathbf{D}_p x[n])$, has an increasingly large eigenvalue. As N increases the global minimum will become sandwiched between two increasingly steep walls— not ideal conditions for gradient descent optimization.

As will be seen in the next section, the waveform error seems well-behaved for *short* chaotic sequences; the exponential amplification of parameter mismatch has little time over which to markedly modify the sequence estimate. Our modification of the waveform error minimization procedure takes advantage of this phenomenon. Instead of trying to optimize our parameter estimates for the whole target sequence at once, we sequentially minimize the waveform errors $E_0, E_1, \ldots, E_N$. Once E_n sinks below some fixed threshold, we repeat the minimization process on E_{n+1}, using the last estimates of the parameters as the initial guess for the next step. Since each error surface has the global minimum in common, successive minimization of the errors forces the parameter estimates closer to their true value. We've effectively expanded the global minimum's basin of attraction to that of a length-one sequence, independent of the true sequence length.

4. EXAMPLES

A simple system capable of exhibiting chaotic behavior is the logistic equation

$$x[n] = p_1 x[n-1](1 - x[n-1]), \quad (11)$$

which is both a scalar dynamical system and a first order nonlinear difference equation. The model exhibits markedly different types of steady state behavior depending on the choice of p_1. A parameter value of $p_1 = 3.5$ results in a period-four oscillation; in contrast, a value $p_1 = 3.7$ produces chaotic behavior. Figure 1 contrasts the waveform error for both cases, for sequences of length $N = 87$ and an initial condition $x[-1] = 0.42$. While relatively flat and smooth in the periodic case, the waveform error in the chaotic case is riddled with local minima and the global

[1] The origin of the sequence can always be shifted to the right by m samples.

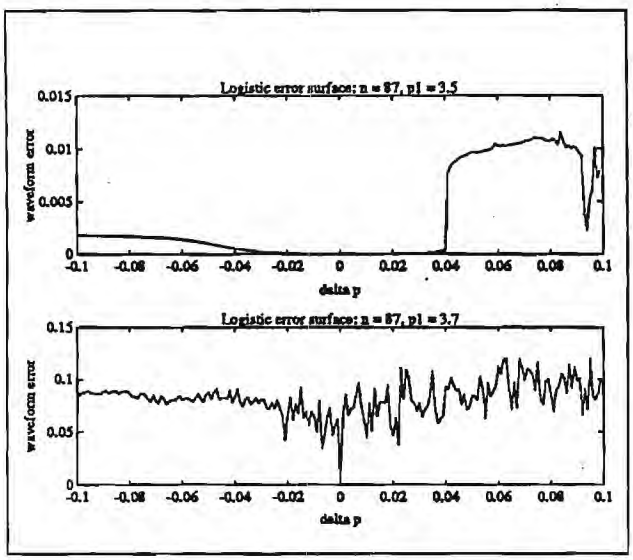

Figure 1: The waveform errors of periodic (above) and chaotic (below) target sequences of length 87, generated by the logistic equation.

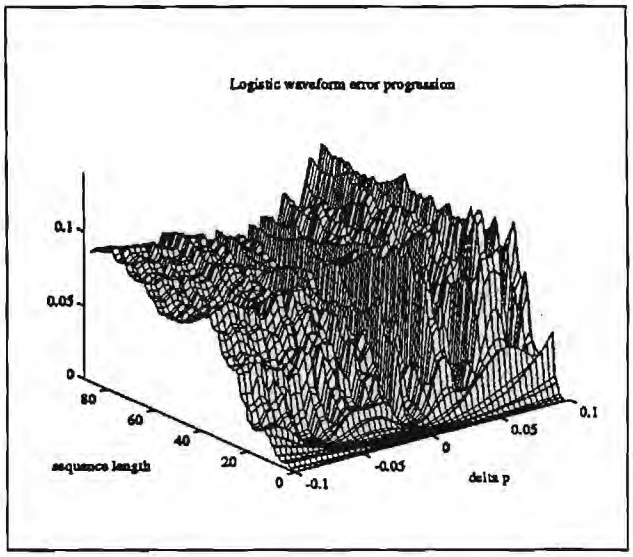

Figure 2: The waveform error as a function of target sequence length. In this case the sequence was generated by the chaotic logistic equation. The error surface is relatively smooth and shallow for short sequences, and becomes increasingly rough as more of the sequence is considered.

minimum has the very narrow basin of attraction as discussed above.

Figure 2 illustrates the behavior of the waveform error for the chaotic logistic system with respect to the sequence length. As noted previously, error surfaces for short sequence lengths are relatively smooth, giving parameter estimates with relatively large errors a greater chance of converging to the true parameter.

Figure 3 compares the performance of the original waveform error minimization technique, which tries to minimize the waveform error E_N directly, and our modified version. As expected from the general appearance of the waveform error, an initial deviation of only 5×10^{-8} in the model parameter gets trapped almost immediately in a local minimum, and the error never descends below the specified threshold of 10^{-5}. Our extension method, on the other hand, correctly identifies the true parameter after extending the target sequence to $N = 87$, even though the initial deviation from the true parameter value was 5×10^{-2}; six orders of magnitudes larger. In fact, any initial parameter value within the logistic equation's usual working range $p_0 \in [0, 4]$ will be converge to the the true parameter value.

An example of a two-parameter chaotic system is the Hénon system

$$y_1[n] = 1 - p_1 y_1^2[n-1] + y_2[n-1] \quad (12)$$

$$y_2[n] = p_2 y_1[n-1] \quad (13)$$

with parameter values $p_1 = 1.4$, $p_2 = 0.3$. If we consider the time sequence produced by the first variable ($y_1[n]$) our difference equation model has the form

$$x[n] = 1 - p_0 x^2[n-1] + p_1 x[n-2], \quad (14)$$

with initial conditions $x[-1] = y_1[-1] = 0.948586$ and $x[-2] = y_2[-1]/p_2 = 0.425317$.

Figure 4 shows that our extension method outperforms the original waveform error minimization technique. Initial parameter estimates with errors of more that 10^{-2} are reduced by six orders of magnitude. The original method, initiated with deviations of only 5×10^{-8} in both parameters, is immediately trapped in a local minimum.

Unlike the one dimensional case, for this two-parameter system our method does not produce estimates that converge to the true parameter values. Figure 5 shows the error surface E_{50} in a small neighborhood of the global minimum. As expected, the sharp gradient discussed in previous sections is in evidence. However, there also seems to be a continuous range of parameter values that generate the same waveform as those located at the global minimum. This alignment is representative of the true behavior of the Hénon system and is not an artifact of the conversion from dynamical system to difference equation. An examination of the gradient of ($y_1[n]$) with respect to the parameters shows that while the ($D_p y_1[n]$) grow quickly for increasing n as expected, they also tend to align themselves along a common axis. The sum of outer products in Eq. 9 is dominated by the matrices formed from these increasingly large gradients that all point in the same direction, and consequently the Hessian appears singular. This behavior is not typical of chaotic dynamical systems in general.

5. CONCLUSION

Using concepts from the discipline of Dynamical Systems we have shown how the sensitive dependence of a chaotic system on its initial conditions can induce an analogous dependence on its parameters. Takens' embedding theorem allowed us to transplant the phase space based notion of Lyapunov exponents which quantify this sensitive depen-

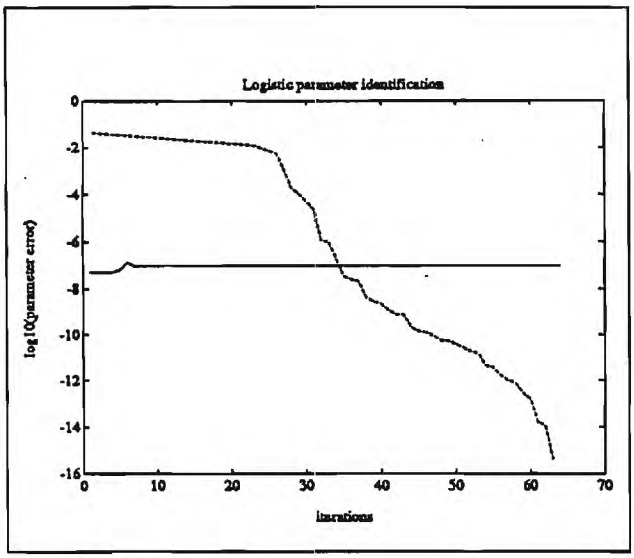

Figure 3: Comparison of waveform error minimization techniques using the chaotic logistic equation. Even good initial parameter estimates get trapped in local minima when trying to minimize the waveform error for the entire target sequence. In contrast, a much poorer initial parameter estimate converges to the true parameter value for a target sequence length of 87 when our modified minimization method is employed.

dence into a nonlinear difference equation framework. We explained how such sensitivity could produce conditions ill-suited for a proposed gradient descent minimization of the waveform error, and proposed an improved method to overcome its limitations. The improved method performed significantly better than the original when tested on sequences generated from two chaotic dynamical systems.

6. REFERENCES

[1] J.-C. Roux, R. H. Simoyi, and H. L. Swinney, "Observation of a strange attractor," *Physica D*, vol. 8, pp. 257–266, 1983.

[2] J. Kurths and H. Herzel, "An attractor in a solar time series," *Physica D*, vol. 25, pp. 165–172, 1987.

[3] J. Gleick, *Chaos*. Sphere Books Limited, 1988.

[4] T. F. Quatieri and E. M. Hofstetter, "Short-time signal representation by nonlinear difference equations," in *Proceedings of the 1990 International Conference on Acoustics, Speech, and Signal Processing (Albuquerque,NM)*, 1990.

[5] J.-P. Eckmann and D. Ruelle, "Ergodic theory of chaos and strange attractors," *Reviews of Modern Physics*, vol. 57, July 1985.

[6] F. Takens, "Detecting strange attractors in turbulence," in *Lectures Notes in Mathematics, vol. 898*, pp. 366–381, Springer, Berlin, 1981.

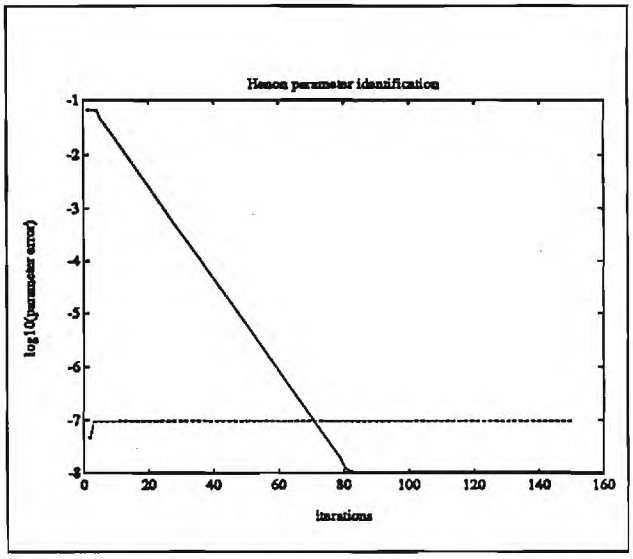

Figure 4: Parameter estimation performance for the Hénon system. Just as in the one-parameter case, the original minimization procedure procedure falls immediately into a local minimum. However, while the modified waveform error method reduces initially much larger parameter deviations better than the original, it does not converge to the true parameters, due to the singular nature of the waveform error's Hessian.

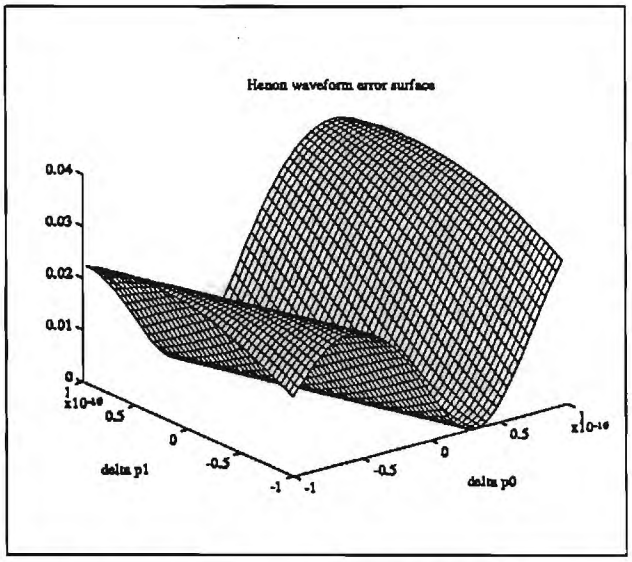

Figure 5: Waveform error surface in the neighborhood of the global minimum. The Hessian appears to be singular.

Submitted to:
ICASSP - 93
Minneapolis, MN, April 1993

Recovering Dynamically Independent Coordinates for Time-delay-embedded Product Systems

Daniel F. Drake and Douglas B. Williams
School of Electrical Engineering
Georgia Institute of Technology
Atlanta, GA 30332
(404) 894-8361

Summary

Takens's time delay embedding result[1] has pushed applications of nonlinear dynamical systems theory beyond the confines of mathematics and physics. The theorem, which proves that a distorted copy of a system's phase space dynamics can be constructed from a sequence of scalar observables, has given researchers from a variety of fields new tools with which to analyze their data: those of nonlinear dynamics. This distorted copy is the product of a nonlinear change of coordinates which preserves certain quantities related to the underlying dynamics—metric entropy, Lyapunov exponents, dimension of the attractor. However, coordinate transformations in general do not preserve the dynamical independence of product systems, i.e. dynamical systems composed of two or more independent subsystems.

The ability to identify dynamically independent coordinates within reconstructed phase space could significantly reduce the complexity of a model derived from a time delay embedding[2]. Formulating the model in the decoupled coordinate system would cause unnecessary cross terms to disappear and provides a natural decomposition into smaller, more easily analyzed independent subsystems.

Product systems arise naturally when considering the problem of separating several unknown, unrelated deterministic signals. The transformation from time delay coordinates to dynamically decoupled axes in the composite waveform's embedding space would allow the contribution of each underlying waveform to be isolated.

We have developed a method to determine if two points in the embedding share a common independent coordinate in the original underlying phase space. We can use this knowledge to construct sets of curves in the embedding which are images of constant-coordinate lines in the original space, thereby creating a new coordinate system that mirrors the independence of the underlying dynamics. While our work is distantly related to that of Fraser[3, 4], who determines the optimal embedding by choosing the time step that maximizes statistical independence of the delay coordinates, a survey of the literature leads us to believe that our approach is quite novel.

For simplicity, imagine two one-dimensional discrete-time dynamical systems $f_1 : x_1[n] \mapsto x_1[n+1]$ and $f_2 : x_2[n] \mapsto x_2[n+1]$ evolving independently. Together they can be viewed as a two-dimensional dynamical system $\mathbf{f} : (x_1[n], x_2[n]) \mapsto (x_1[n+1], x_2[n+1])$ with dynamically independent coordinates—a product system. Now suppose we apply a simple change of coordinates which skews the x_2 coordinate with respect to x_1 (see figure 1). In the new coordinate

system (y_1, y_2), two points sharing a common y_1 coordinate value no longer map via the dynamics to other points with identical y_1 components, i.e. the new y_1 coordinates are dependent on y_2 values. Coordinate transformations do not preserve dynamical independence.

In general, the transformation associated with the time delay embedding technique is more complex than a simple skew (see figure 2). Suppose the dynamical system $\mathbf{f}$ generates a scalar observable via $g : (x_1[n], x_2[n]) \mapsto y[n]$. Then for a suitably large embedding dimension m and under certain loose conditions on $\mathbf{f}$ and g, Takens proved that an invertible change of coordinates relates the set of vectors

$$\mathbf{y}[n] = (y[n], y[n+1], \ldots, y[n+m-1])$$

to the points $\mathbf{x}[n] = (x_1[n], x_2[n])$ of the original phase space.

Our method exploits the implicit dependence of the embedding on the underlying dynamics. Besides the assumptions that

1. the conditions set forth by Takens are met;
2. the underlying dynamics are decoupled;
3. $$\frac{\partial^2 g(x_1, x_2)}{\partial x_1 \partial x_2} = \frac{\partial^2 g(x_1, x_2)}{\partial x_2 \partial x_1} = 0;$$

no further knowledge of the underlying system is necessary. The strong sensitivity that accompanies chaotic dynamics can be exploited to refine the precision of the constant-coordinate curves. Concerning assumption 2, we are currently investigating the possibility of testing for independence before applying our method, thereby avoiding spurious results. Assumption 3 allows combinations like $g(x_1, x_2) = e^{x_1} + \ln x_2$, but the modulation $g(x_1, x_2) = x_1 x_2$ will not be interpreted correctly (though we could work with the logarithm of the time series). We hope to relax this restriction in the future.

Other submissions to ICASSP-93 with the same co-author

Soonjoo Hwang and Douglas B. Williams: A Total Least Squares Approach for Array Processing with Unknown Sensor Positions

Wayne T. Padgett and Douglas B. Williams: Time-Delay Estimation for Damped Sinusoids Incident on an Array

References

[1] F. Takens, "Detecting strange attractors in turbulence," in *Lecture Notes in Mathematics, vol. 898*, pp. 366–381, Springer, Berlin, 1981.

[2] J. P. Crutchfield and B. S. McNamara, "Equations of motion from a data series," *Complex Systems*, vol. 1, pp. 417–452, 1987.

[3] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Physical Review A*, vol. 33, pp. 1134–1140, February 1986.

[4] A. M. Fraser, "Information and entropy in strange attractors," *IEEE Transactions on Information Theory*, vol. 35, pp. 245–262, March 1989.

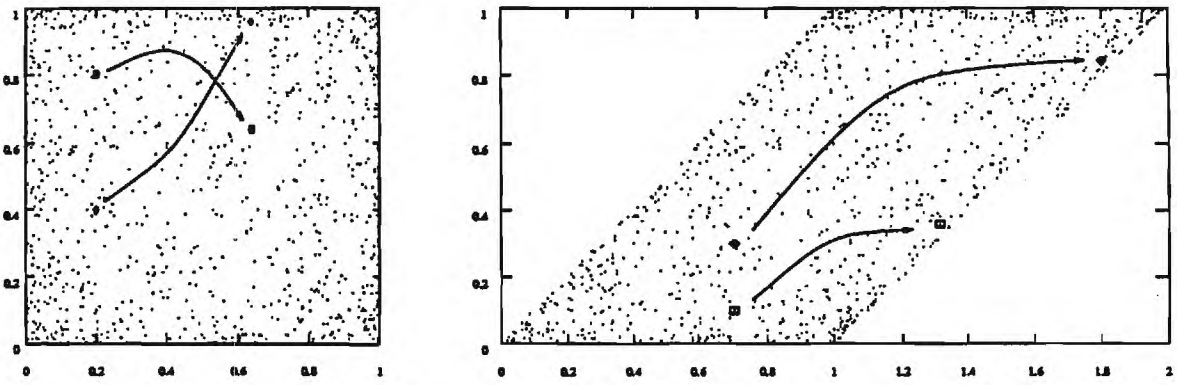

Figure 1: Left: a product system composed of two one-dimensional dynamical systems. The dynamics along each axis are governed independently by the logistic equation $x[n+1] = 4x[n](1-x[n])$. Points with identical $x_1[n]$ coordinates are mapped by the dynamics to points with common $x_1[n+1]$, independently of $x_2[n]$. The small dots represent a trajectory of the system; they serve to delineate the phase space boundaries. Right: A skewed coordinate transformation $T : (x_1, x_2) \mapsto (y_1, y_2)$ eliminates the dynamical independence by coupling $y_1[n+1]$ to $y_2[n]$.

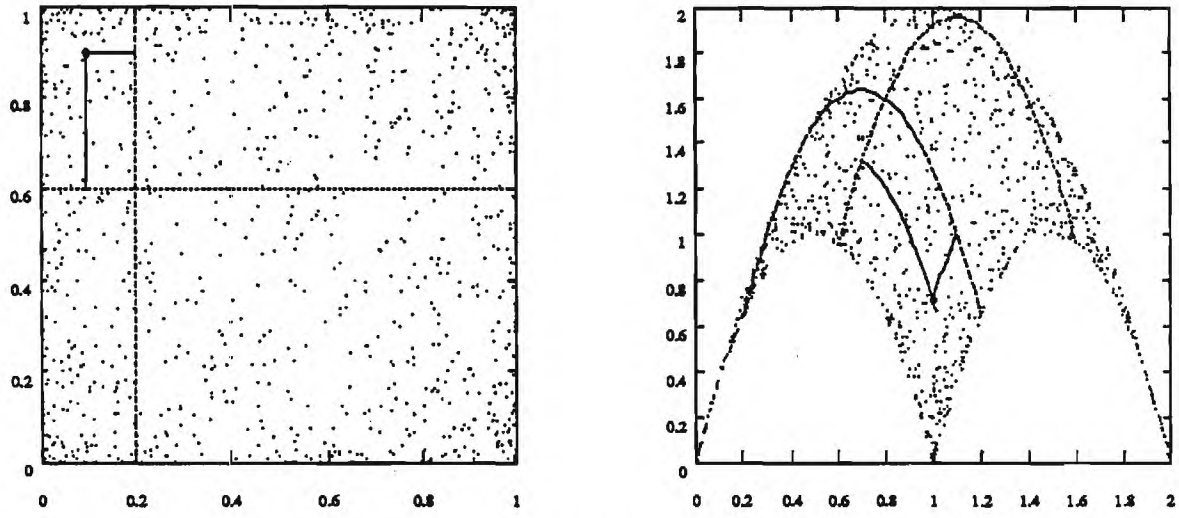

Figure 2: Left: the same product system as in the previous figure. The lines $x_1 = 0.2$ and $x_2 = 0.6$ represent a set of "constant-coordinate" lines along which one component of the dynamics is fixed. The point at $(0.1, 0.9)$ has been projected onto the constant-coordinate lines. Right: A time delay embedding transformation of the product system on the left for observable $y[n] = x_1[n] + x_2[n]$. The images of the constant-coordinate lines form axes for the decoupled coordinate system in the embedding space. The image of the point $(0.1, 0.9)$ is shown "projected" onto the new axes along constant-coordinate curves (images of translates of the constant coordinate lines in the underlying system).